The Coding of Sound Localization Cues in the Inferior Colliculus

by

Steven M. Chase

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

June 2006

Abstract

The auditory system uses three cues to decode sound location: interaural time differences (ITDs), interaural level differences (ILDs), and spectral notches (SNs). Initial processing of these cues is performed in several auditory brainstem nuclei that send projections to neurons of the inferior colliculus (IC). This work addresses how information about these different sound localization cues is integrated into the responses of single neurons of the IC.

Virtual space techniques were used to create stimulus sets varying in two sound-localization parameters each. By manipulating pairs of cues within a stimulus set, the relative coding of each cue could be compared. Using a variety of information theoretic methods, the mutual information between the localization cues and the neural response was quantified under the assumption of several different encoding schemes.

The results show that the three localization cues are best represented by different codes. ITD information is conveyed by spike rate alone, and is contained only in low frequency neurons. ILD information is best represented by a joint rate/first spike latency code. The coding of SNs changes with the best frequency (BF) of the neuron. Low BF neurons represent SNs by the timing of spikes distributed throughout the response, where the spike times are locked to particular stimulus features. High BF neurons, on the other hand, represent SNs by spike rate and, to a lesser extent, first spike latency. The differential coding of the localization cues suggests that information about multiple cues could be multiplexed onto the responses of single neurons.

These results have implications for how the localization cues might be integrated into a percept of sound location. The accuracy with which each cue can contribute to the overall location percept changes depending on sound conditions, such as the frequency content of the stimulus, temporal characteristics of the stimulus, or the reverberant qualities of the environment. That the cues are differentially coded in the IC implies that the brain may have access to the individual cues at cortical levels, and the weight with which each cue contributes to the location percept could be tailored to the stimulus, environment, and perceptual task.

Advisor

Dr. Eric D. Young

Professor

Biomedical Engineering

Acknowledgements

Of course, I could never have completed this work on my own. My gratitude for the people whose patience, understanding, and insight helped me through this dissertation cannot adequately be expressed in words. I am indebted.

My advisor, Eric Young, for giving me the guidance to get me going and the freedom to let me find my own way.

Brad May, my second reader, who was always willing to answer questions and discuss results.

Dissertation committee: Ed Connor, Xiaoqin Wang, and Kechen Zhang.

NEL members, past and present, without whom this would have been a dreary and caffeine poor experience: Mike Anderson, Sharba Bandyopadhyay, Ian Bruce, Shanqing Cai, Kevin Davis, Gulam Emadi, Alon Fishbach, Mike Heinz, Ben Letham, Diana Ma, Lina Reiss, Tessa Ropp, Danilo Scepanovic, Zach Smith, William Tam, Phyllis Taylor, Anita Tilotta, Josh Vogelstein, and Jane Yu.

My friends and family, who never admitted boredom when listening to the last result or setback.

And finally, Anita Zuberi, who has always been a bottomless well of love and support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

1. Introduction and Background

In general, the mechanisms of perception are similar for all modalities: the sensory stimulus is transduced into a series of electrical impulses, spikes, that are conveyed by an array of neurons to specific brain regions. If a particular stimulus feature is to be perceived, it must somehow by encoded in the timing (and numbers) of these spikes across the neural substrate. In a few cases, such as the retina, and to a lesser extent the auditory nerve, the transformation from stimulus to spike is fairly well understood (Pillow *et al.*, 2005; Zhang *et al.*, 2001), in the sense that the probabilities of single spike times can be predicted with high accuracy. As higher levels, however, information is split and recombined in such a way as to extract higher-level features that are relevant to the organism, and the relationship between spikes and stimulus features becomes blurred. Outside of the periphery, the mechanisms by which spikes represent stimulus features are not well understood.

This dissertation is an attempt to address this general problem in a specific, well-defined setting. Here, the representation of sound localization cues by single neurons of the central nucleus of the inferior colliculus (IC) is investigated. By focusing attention on

this very specific problem, it is hoped not only to understand how neurons extract and process information about the location of a sound source, but also to come up with a methodology appropriate for investigating general relationships between spikes and stimulus parameters.

The general approach is as follows. Using virtual space techniques, several stimulus sets are created that vary in two sound localization cues each. Each individual cue varies in 5 steps over a physiologically meaningful range, and all possible combinations of the two cues are taken, for a total of 25 stimuli per set. The idea is to test the sensitivity of each neuron to changes in one localization cue versus the other. To do this, the 25 stimuli are presented multiple times to build up statistics sufficient to describe the probability of some feature of the neural response, such as rate or first spike latency, to each combination of localization cues. Information theory is then used to quantify the sensitivity of the neuron to manipulations in the cues.

This approach allows one to address several questions about the nature of the coding of sound localization cues. Do different physiological neuron types respond preferentially to certain cues over others? Are different cues represented by different codes? Are cues encoded in such a way that higher brain centers have access to each individual cue, independently of the other? These questions will be formally addressed in subsequent chapters.

This chapter proceeds with a discussion of sound localization cues and a review of the current understanding of their processing in structures up to and including the IC. Chapter 2 reviews the construction of the stimuli, operating procedures, and recording methods. In Chapt. 3, mutual information (MI) is defined, and several methods for

estimating it are contrasted. The next three chapters are devoted to results under the assumption of various codes, including rate (Chapt. 4), temporal decoders (Chapt. 5), and first spike latency (Chapt. 6). Finally, general conclusions are summarized in Chapt. 7.

1.1 Cues for sound localization

There are three main cues for sound localization (for review, see Middlebrooks and Green, 1991). When a sound source is located to the left of a listener's head, it will arrive at the left ear before arriving at the right ear. This interaural timing difference (ITD) varies in a systematic manner with the azimuth, and data from psychophysics shows that ITDs provide a strong cue for azimuth at low frequencies (Mills, 1958; Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002). The ultimate dynamic range of the cue varies with the size of the head; in humans, physiological ITDs vary from approximately -700 μs to 700 μs; in cats the range is from approximately –350 μs to 350 μs (Roth *et al.*, 1980).

Sounds on the left side of a listener will also be louder in the left ear than they are in the right, due to the shadowing effect of the head. While the exact value of this interaural level difference (ILD) varies with frequency due to the filtering of the pinnae, to first order it can be modeled as an attenuation applied equally to all frequencies. Data from psychophysics also shows that ILDs provide a strong cue for azimuth, especially at high frequencies (Mills, 1958; Macpherson and Middlebrooks, 2002).

Finally, the bumps and folds of the pinnae filter the incoming sound wave in a location-dependent manner. The head-related transfer function (HRTF) characterizes this

filtering process as the gain (in dB) of the signal at the eardrum relative to the signal at the source, as a function of frequency and source location. Some examples of these HRTFs are shown in Fig. 1.1, from the cat data of Rice *et al.* (1992). The most prominent feature of the HRTF that changes with location is the first spectral notch (SN), located in the example figure between 9 and 12 kHz. The center frequency of this notch changes with the location of the sound in a systematic manner; in cats, the frequency range spanned by these SNs is from 6 to 20 kHz (Musicant *et al.*, 1990; Rice *et al.*, 1992). Errors in the judgment of source elevation result when noises are presented to cats with this frequency range removed (Huang and May, 1996), suggesting that these SNs represent a cue for some aspects of sound localization.

1.2 Anatomy and physiology sound localization nuclei

A schematic of the major circuitry in the mammalian auditory brainstem leading up to the IC is given in Fig. 1.2. A number of anatomical specializations have evolved to process sound localization cues; the major pathways by which sound localization information reaches the IC are briefly described below. For a review of this topic, and auditory brainstem anatomy and physiology in general, see Irvine (1992), Oliver and Huerta (1992), and Webster *et al.* (1992).

1.2.1 The ITD processing pathway

The ITD processing pathway, outlined in green in Fig. 1.2, starts with the auditory nerve. When low frequency auditory nerve fibers (ANFs) receive tonal input, they will preferentially fire at a certain phase of the tone, a phenomenon known as phase-locking (Rose *et al.*, 1967). When ANFs receive broad-band input, they tend to phase-lock to the frequency components nearest their best frequency (BF, describing the frequency of the lowest level tone that elicits a response) (deBoer and deJongh, 1978). These ANFs synapse, among other places, on the spherical bushy cells of the anteroventral cochlear nucleus (AVCN; for review see Cant, 1992).

Spherical bushy cells, which receive strong synaptic input from a small number of ANFs of similar BF (Liberman, 1991; Ryugo and Sento, 1991; Rothman *et al.*, 1993), possess low threshold potassium channels that give them a very fast membrane time constant (Manis and Marx 1991; Wu and Oertel, 1984). The fast membrane time constant and strong synaptic input allows these neurons to preserve the precise phase-locking that exists in low frequency ANFs; indeed, some studies suggest that the phase-locking may be tighter in bushy cells than it is in the auditory nerve (Joris *et al.*, 1994; Louage *et al.*, 2005).

ITDs are first compared in the medial superior olive (MSO), which receives convergent, excitatory input from low frequency spherical bushy on both sides of the head (Stotler, 1953). Because it is difficult to obtain clean electrophysiological recordings from MSO cells, there is not a lot of data on their responses to sound. However, the studies that do exist have shown that these cells exhibit clear tuning to ITDs, with the highest firing rates occurring at the cell's characteristic delay (Goldberg and Brown, 1969; Yin and Chan, 1990). The initial hypothesis, formulated by Jefress

(1948), postulated that this ITD tuning was formed by synaptic delays arising from the travel time differences of action potentials from the two AVCN. More recent data, however, suggests that inhibition likely plays a role in shaping the ITD tuning of MSO cells (Brand *et al.*, 2002). Information about ITDs reaches the IC both directly from projections of the MSO, and indirectly via the dorsal nucleus of the lateral lemniscus.

1.2.2 The ILD processing pathway

The predominant pathway by which ILD information ascends to the IC, outlined in red in the schematic of Fig. 1.2, is via the lateral superior olive (LSO). Like the MSO, the LSO receives excitatory input from the ipsilateral AVCN. However, unlike the MSO, contralateral input to the LSO is inhibitory. Globular bushy cells (GBCs) of the AVCN receive input from, typically, about 20 ANFs (Spirou *et al.*, 2005). These cells project to the medial nucleus of the trapezoid body (MNTB), which contains inhibitory neurons that project to the contralateral LSO. By receiving excitatory input ipsilaterally and inhibitory input contralaterally, LSO cells are ideally situated to process ILDs, and physiological studies of these neurons reveal a strong ILD sensitivity (Boudreau and Tsuchitani, 1968; Guinan *et al.*, 1972a,b; Tollin and Yin, 2002). Cells in the LSO project directly to the IC.

1.2.3 The SN processing pathway

The predominant pathway for the processing of SN information as it pertains to sound localization is via the dorsal cochlear nucleus (DCN), as illustrated in blue in Fig. 1.2.

The principal cells of the DCN receive direct input from ANFs as well as both local and non-local inhibitory input (for review, see Young and Davis, 2002). The circuitry is such that DCN principal cells show firing rate changes to monaural SNs positioned near their BF (Young *et al.*, 1992; Imig *et al.*, 2000; Reiss and Young, 2005). Furthermore, psychophysical studies performed in cats with DCN lesions have found a decreased ability to localize sounds, especially in elevation (Sutherland, 1998; May, 2000). DCN principal cells project directly to the contralateral IC.

1.2.4 Overlaps in the localization cue pathways

The segregation of localization cue pathways described above is not complete. MSO neurons respond strongly to ITDs, however LSO neurons also show a weak response to ITDs (Joris and Yin, 1995). While LSO neurons probably do the majority of ILD-related processing (Tollin, 2003), MSO neurons also respond weakly to ILDs (Goldberg and Brown, 1969), as do DCN neurons (Young and Brownell, 1976; Joris and Smith 1998), and ILD-related processing could potentially be computed from monaural streams converging at the level of the NLL or even in the IC itself. Furthermore, neurons in both the MSO and LSO will show a decrease in firing rate when a SN is centered on their receptive fields, as would any narrowly tuned neuron.

These caveats aside, it is likely that the three localization cues activate predominantly different circuits in the auditory brainstem, as illustrated in Fig. 1.2. The manner in which these cues convergence to form a coherent percept of spatial location is still not well understood. As an early site of convergence for information about the three cues,

the IC represents an ideal place to probe how information from the different streams might be put together.

1.3 Previous studies of the inferior colliculus relevant to sound localization

The IC is tonotopically organized, possessing a frequency axis that runs dorsoventrally from low to high BF in cats (Merzenich and Reid, 1974). As a major stop in the transfer of ascending acoustic information and the last processing point in the acoustic brainstem, the IC receives convergent input from nearly every lower acoustic nucleus (Nordeen *et al.*, 1983). Anatomical studies show that inputs from different brainstem nuclei appear to form a partly overlapping map, potentially producing regional variations in convergence within the IC (Roth *et al.*, 1978; Brunso-Bechtold *et al.*, 1981; Oliver *et al.*, 1997). This convergence allows cells in the IC to be sensitive to binaural location cues embedded in virtual space stimuli (Delgutte *et al.*, 1995; Delgutte *et al.*, 1999).

1.3.1 Physiological cell types within the inferior colliculus

Ramachandran *et al.* (1999) defined four classes of neurons in the IC of decerebrate cats based on responses to tone bursts. These classes are known as type-V neurons, type-I neurons, type-O neurons, and onset neurons (Fig. 1.3). Type-V neurons are found only in the low frequency region of the nucleus (< 5 kHz) and are excited by both ipsilateral and contralateral tones. They are sensitive to ITD information carried by tones

(Ramachandran and May, 2002) and show a sensitivity to average binaural intensity (ABI) (Davis *et al.*, 1999). Since the behavior of these neurons mimics that of neurons in the MSO (Yin and Chan, 1990), Ramachandran and colleagues hypothesized that the MSO is the dominant source of input for this class of neurons.

Type-I neurons are found in all frequency regions, are excited by contralateral tones in a narrow spectral region around their BF, and typically show inhibition to tones further from BF or presented ipsilaterally. As type-I neurons are sensitive to ILD (Davis *et al.*, 1999) and show some sensitivity to envelope ITDs (Ramachandran and May, 2002), it was hypothesized that the LSO is the dominant input to this class of cells.

The type-O neuron is also found in all frequency regions, is excited by contralateral BF tones, and typically is inhibited by ipsilateral tones (Davis *et al.*, 1999). It has a response field that looks much like that of a DCN type-IV neuron (Young and Davis, 2002), and pharmacological evidence suggests that the dominant input to type-O neurons arises from this source (Davis, 2002). Furthermore, like the DCN (Reiss and Young, 2005), type-O neurons have been shown to be sensitive to spectral notches (Davis *et al.*, 2003).

Onset neurons are rare, comprising less than 10% of the total population in unanesthetized animals; they respond with a single onset spike to tone bursts over a broad range of frequencies and levels.

1.3.2 Sound location processing within the inferior colliculus

Neural correlates of various psychophysical phenomena involving localization cues have been found at the level of the IC. Low frequency IC neurons sensitive to ITDs have been shown to exhibit sensitivity to binaural masking level differences (Licklider, 1948), where a tone masked by noise can be made audible by changing the phase of either the tone or the noise in one ear (Jiang *et al.*, 1997; Palmer *et al.*, 2000). Also, neural correlates of the precedence effect have been found in the IC (Yin, 1994; Litovsky and Delgutte, 2002).

The sensitivity of mammalian IC neurons to sound localization cues has been extensively studied. Only a few studies, however, have looked at the effects of multiple cues simultaneously. Semple and Kitzes (1987) examined the interaction of ABI and ILD sensitivity in the IC of gerbils using pure tone stimuli, as did Irvine and Gago (1990) in cats and Wenstrup *et al.* (1988) in bats. The interaction of ITD and ILD sensitivity in the IC of cats has been investigated with clicks (Benevento and Coleman, 1970) and tones and broadband noise (Caird and Klinke, 1987). Delgutte *et al.* (1995) manipulated virtual space stimuli directly to selectively target the effects of ITD and ILD cues on azimuth tuning. In general, these studies have found populations of neurons dominated by ILD cues, ITD cues, both cues, or neither cue, as assessed by spike rate changes.

A few studies have shown that spike timing can also carry information about sound localization cues in the auditory system. Middlebrooks and colleagues (1994) showed that an artificial neural network could be trained to use spike trains from single neurons in the auditory cortex to discriminate sounds from different locations; this discrimination was greater given full spike trains than just spike counts alone. Subsequently, they discovered that the majority of information was transmitted by the first spike latency

(Furukawa and Middlebrooks, 2002). First spike latency modulations have also been observed in the IC in response to sound location changes (Sterbing *et al.*, 2003; Behrend *et al.*, 2004).

1.4 Characterizing the responses of auditory neurons

For years, auditory neurons have been characterized by responses to fairly simple stimuli such as pure tones and broadband noise on the basis of tuning curves, rate-level functions, post-stimulus time histograms (Kiang *et al.*, 1967), and reverse correlation methods (Eggermont *et al.*, 1983). While these measures are good for the initial characterization, the non-linear nature of most auditory neurons makes these descriptions incomplete. For example, it is impossible to predict the response of a Type IV DCN neuron to broadband noise based on its responses to pure tones (Spirou and Young, 1991; Nelken *et al.*, 1997).

More recently, other measures have been developed for more general categorizations of auditory neurons. One such method is with a spectral weighting function, where an input-output function relating neural firing rate to the stimulus amplitude of certain frequency components is estimated and used to predict the responses to arbitrary stimuli (Yu and Young, 2000). This method considers only steady-state spectral shape and does not consider the temporal sensitivity of neurons. Another method is via spectral temporal receptive fields, which are linear transfer functions relating neural firing rate to the stimulus amplitude as a function of frequency and time (Aertsen and Johannesma, 1981). These have been calculated using reverse correlation methods, dynamic ripple spectra

stimuli (Depireux *et al.*, 2001), or even naturalistic stimuli (Theunissen, 2000). While these methods can be used to provide insight into the behavior of even very complex neurons, in general they are limited to linear or low order non-linear dependencies due to the added computational and experimental difficulties. For this reason, they turn out to be inadequate to fully describe the behavior of more complex neurons. These methods are also limited in that they make assumptions about the encoding mechanisms. Usually, it is assumed that either each spike is independent of every other spike or that the relevant information is contained entirely in a particular neural rate code, an assumption that in many cases is not correct (de Ruyter van Steveninck and Bialek, 1988).

In this study, information theory is used to quantify the sensitivity of the neural response to changes in stimulus parameters. With mutual information (Shannon, 1948; Cover and Thomas, 1991), the neural response to auditory stimuli can be characterized with a minimum number of assumptions about how the stimulus information is encoded, taking both temporal and rate codes into account. Since the calculations depend only upon the probability distributions of spike times in response to stimuli, this method also makes no assumptions about whether the encoding is done linearly or non-linearly (Johnson *et al.*, 2001; Strong *et al.*, 1998). In this sense, information theoretic measures are ideal for characterizing how well neurons respond to various stimulus parameters.

The results show that the three localization cues are differentially encoded in the neural response. ITD information is conveyed by spike rate only, and is conveyed by a subset of low BF neurons of each physiological type. ILD information is best represented by a joint rate/first spike latency code, and is strongly encoded in all cell types except type-V. The coding of SNs changes with the best frequency (BF) of the

neuron. Low BF neurons of all cell types represent SNs by the timing of spikes distributed throughout the response, where the spike times are locked to particular stimulus features. Spike rate is not an effective code for SN in low BF neurons. High BF neurons, on the other hand, represent SNs by spike rate and, to a lesser extent, first spike latency. The differential coding of the localization cues suggests that information about multiple cues could be multiplexed onto the responses of single neurons.

2. Methods

This chapter provides an overview of the surgical procedure (section 2.1), electrode fabrication process (section 2.2), recording procedure (section 2.3), and stimulus construction process (section 2.4).

2.1 Surgical procedure

Experiments were performed on adult cats with clean external ears and middle ears free from infection. Cats were typically about 2-3 kg in weight. A surgical plane of anesthesia was achieved with xylazine (1 mg/kg IM) and ketamine (40 mg/kg IM), maintained with supplemental doses (~15 mg/kg IV of ketamine) as needed. At the beginning of surgery cats were given atropine sulfate (0.1 mg IM) to reduce mucous secretions and dexamethasone (2.0 mg IM) to reduce edema.

Once anesthetized, cats were prepped for surgery by shaving the top of the head, forelimbs, neck, and base of the tail. A rectal thermometer was used to maintain the cat's temperature between 37.5 °C and 38.5 °C with a feedback-controlled heating pad. The

cephalic vein was then cannulated to allow the IV perfusion of drugs and fluids during the course of the experiment and a tracheotomy was performed to facilitate breathing.

The main surgery began with an incision along the midline of the scalp, from just above the orbital ridge to just above the top of the spine. Typically, a Y-incision was made around the top of the spine to reduce tension. The temporal muscle was then reflected on one side of the scalp to provide access to the parietal bone. To expose the brain for decerebration, a hole was opened through the parietal bone approximately 1 cm rostral to the interaural axis, at an angle in the frontal plane of ~40º to the top of the skull. The cat was decerebrated by aspirating the brain matter between the superior colliculus and the thalamus, completely severing the thalamus and cortex from the brain stem. Once the decerebration was complete, anesthesia was discontinued.

Next, the ear canals were exposed on both sides by fully reflecting the skin and temporal muscles and clearing away the fascia. The ear canals were then partially transected approximately half a centimeter from their entrance into the skull and fitted with hollow brass ear tubes for sound delivery. The cat was fit into a stereotaxic frame connected to the ear bars and stabilized with a clamp around the orbital sockets and roof of the mouth. To prevent pressure buildup in the middle ear, the bullae on both sides were exposed and vented with ~30 cm of polyethylene (PE 90) tubing.

The IC was exposed dorsally by drilling through the skull overlying the occipital cortex and aspirating cortex until the IC was visualized. In most cases, the bony part of the overlying tentorium was removed and the overlying dura reflected to gain full access to the IC. At the end of the experiment, the cat was killed with an overdose of barbiturate anesthetic (975 mg to effect). All procedures were performed in accordance with the

guidelines of the Institutional Animal Care and Use Committee of the Johns Hopkins University.

2.2 Electrode fabrication

In-house fabricated platinum/iridium microelectrodes were used for single neuron recordings. Platinum/iridium (70/30) wire of 0.004″ or 0.005″ diameter was obtained from the supplier in spools. As the wire typically takes on the curvature of the spool, it requires straightening before use. Wire was straightened by clamping it in series with a current source while suspending weight from it. A current was then applied at levels high enough to heat the wire to the glowing point. Straightened wire was sectioned into segments of ~2 cm in length for electrodes.

Each section of straightened wire was clamped into 2.5″ of stainless steel tubing with an outer diameter of 0.014″ and an inner diameter just greater than 0.005″. The electrode shank was chemically-etched by passing current through it while immersing it in NaCN solution (buffered with NaOH). The wire was sharpened to a point by using a motor to slowly move the electrode in and out of solution while lowering the voltage between dips. The sharpness of the tip depends on various dipping parameters such as the number of dips, dip speed, voltage, immersion depth, and the freshness of the NaCN solution; the parameters were adjusted to achieve a tip slope of approximately 4:1.

Once sharpened, the electrode was insulated with a layer of glass applied by melting glass powder over the tip and shank, up to the joint with the stainless steel tubing. A

glass capillary tube was fitted over the tubing and glued to the melted glass at the bottom, leaving ~1 cm of tubing sticking out of the top for connecting to the recording amplifiers.

Finally, a platinum ball was applied to the very tip of the electrode by applying a voltage (~10 V) to the electrode while touching the surface of a platinum/chloride solution (0.5% H_2PtCl_6) connected in series. Touching the tip to the surface completes the circuit and causes the current to short through the melted glass at the tip of the wire, leading to deposition of a ball of platinum. Typical tips achieved a hemispherical shape approximately 5 μms in diameter and had a resistance of 0.5 to 2.0 MΩ.

2.3 Recording procedure

Recordings were made in a sound-attenuating chamber, with sounds presented via speakers placed on the free end of the hollow ear bars. The acoustical system was calibrated *in situ* for both ears with a probe tube microphone placed ~2 mm from the cat's tympanic membranes. This system produces a fairly uniform response between 40 Hz and 35 kHz (±4.6 dB SPL).

The electrode signal was amplified 10,000-30,000 times and bandpass filtered between 100 Hz and 6 kHz. Single neurons were isolated with a Schmitt trigger or a template-matching program (Alpha-Omega Engineering, Israel). All data are based on clear single neuron recordings. The BFs of neurons were determined manually.

Electrodes were advanced dorsoventrally through the IC to sample neurons with various BFs. Electrode entry into the central nucleus was determined by physiological cues as follows. As the electrode was advanced, BFs first decreased, then increased.

This frequency reversal occurred ~1-2 mm below the surface and signified transition from dorsal cortex or external nucleus into the central nucleus (Rose *et al.*, 1963; Merzenich and Reid, 1974; Aitkin *et al.*, 1975). Neurons in the dorsal cortex or external nucleus were often tuned poorly, adapted rapidly, or exhibited marked offset responses; neurons within the ICC were usually sharply tuned with short latencies (Merzenich and Reid, 1974; Aitkin *et al.*, 1975). The ventral edge of the central IC was reached when the BF jumped abruptly from high frequencies (~30-40 kHz) to low frequencies (<1 kHz), signifying entry into the dorsal nucleus of the lateral lemniscus (Aitkin *et al.*, 1970), or when background activity to auditory stimulation disappeared.

Once a neuron was isolated, characterization stimuli were presented to compare the neuron to known response types of IC neurons, as in Fig. 1.3. All characterization stimuli were 200 ms in length including 10 ms linear rise/fall edges, followed by an 800 ms silent interval. The average discharge rate over the whole 200 ms interval was used for the classification; Ramachandran and colleagues (1999) used the last 150 ms of the stimulus for their rate calculations; the response classifications are robust and do not change if rates are computed from 150 or 200 ms. If rates are computed over shorter intervals (say the first 50 ms), inhibitory regions become harder to see, so classification becomes more difficult. Spontaneous rates were calculated over the last 400 ms of the silent period.

Classification was done as follows. Neurons that were excited by monaural tones presented to either ear and that had little inhibition in the response map were classified as type-V. Neurons whose responses to contralateral BF tones were excitatory at low levels but turned inhibitory at high sound levels were classified as type-O. Neurons that were

excited at all contralateral BF tone levels and displayed clear sideband inhibition were classified as type-I. The majority of these neurons were also inhibited by ipsilaterally presented tones, although there were 3 high BF neurons classified as type I that were binaurally excited with clear sideband inhibition. Consistent with Ramachandran, et al. (1999), no type-V neurons were found with BFs above 4.8 kHz. The classification of low BF (<3 kHz) neurons is difficult. Qualitatively, while the distinction between physiological classes is quite clear in the mid-frequency region (3-20 kHz), below 3 kHz the class boundaries become fuzzy. Almost all the low-BF neurons in our sample had type-V characteristics.

Once the neuron was characterized as type V, I, or O, analysis stimuli (described below) were presented for the remainder of the recording time. The stimulus level was chosen to be near the center of the dynamic range of the neuron under study.

2.4 Stimulus design

Three sets of virtual space stimuli were constructed. Each stimulus set consisted of a 330 ms broadband frozen noise token (sampled at 100 kHz, interstimulus interval 1 s) that was manipulated to vary independently over 2 parameters. Each parameter was adjusted in 5 steps, for a total of 25 stimuli per set. To build up statistics sufficient for information theoretic analyses, each stimulus set was repeated from 20-200 times, depending on how long contact with the neuron was maintained (median: 70 repetitions). These 25 stimuli were presented in fixed order, and the individual stimuli were

interleaved; that is, all 25 stimuli were presented once, then the whole stimulus set was repeated, etc.

The first stimulus set, shown in Fig. 2.1A, was designed to probe the differential sensitivity to ITD and ILD cues. The noise token was filtered through a spatially averaged head related transfer function (HRTF); the HRTFs used here were obtained from the cat data of Rice *et al.* (1992), and represent the region in space from –60 to 60 degrees azimuth and –30 to 45 degrees elevation. Filtering the noise token through a spatially averaged HRTF imparts to the stimulus the spectral characteristics of the head and ear canal, independent of spatial location. The stimulus was then split into two streams (one for each ear) that were delayed relative to one another to impart an ITD. Five ITDs were used, at -160, -80, 0, 80, and 160 μs, where a positive value indicates the contralateral stimulus is leading. Physiological ITD-azimuth functions vary with frequency, elevation, and head size, and depend upon whether transient or steady-state values are measured (Kuhn, 1977; Roth *et al.*, 1980). The values chosen here correspond roughly to spatial locations in the horizontal plane of -60, -30, 0, 30, and 60 degrees azimuth, on a hard sphere with the approximate diameter of a cat's head; transient measurements for clicks yield somewhat longer delays for the same azimuths (Roth *et al.*, 1980). Finally, the two streams were attenuated relative to one another to impart an ILD. Five ILDs were used, at -13.8, -8.4, 0, 8.4, and 13.8 dB. When averaging the ILDs measured from cat HRTF functions over frequency and elevation, these values also correspond roughly to -60, -30, 0, 30, and 60 degrees azimuth. The two most popular methods of imposing ILDs in the literature are the contra-constant method, where the contralateral level is held fixed while ILD is manipulated with the level in the ipsilateral

ear (Rose *et al.*, 1966; Geisler *et al.*, 1969; Davis *et al.*, 1999; Park *et al.*, 2004), and the average binaural intensity (ABI) constant method, where the average summed level is held constant, and the levels in both ears are adjusted to create the ILD (Phillips and Irvine, 1981; Wise and Irvine, 1985; Semple and Kitzes, 1987). In the first two stimulus sets, a variation of these methods was used. For the 5 ILDs, the levels in the ipsilateral ear were set to 0, 0, 0, -8.4, and -13.8 dB (re the reference level), and the levels in the contralateral ear were set to -13.8, -8.4, 0, 0, and 0 dB. This non-standard method of ILD adjustment was originally done so that, in a plot of response rate vs. ILD, the effects of ILD could be dissociated from the effects of monaural response to either contralateral or ipsilateral stimulation, something that cannot be done with either the contra-constant or the ABI constant methods. The third stimulus set (described below) was designed to disambiguate binaural and monaural responses more directly.

The second stimulus set was designed to probe the differential sensitivity to ILD and SN cues (Fig. 2.1B). The SN cue was imparted by filtering the frozen noise token through one of five HRTFs containing a prominent spectral notch. These HRTFs, shown in Fig. 1.1, were taken from the azimuth midline and correspond to 0, 7.5, 15, 22.5, and 30 degrees elevation. The stimulus was then split into two streams, and an ILD cue was imparted as in the ITD/ILD stimulus set. Note that since the stream splitting occurs after filtering, the same spectral shape is sent to each ear. Interaural spectral differences are not considered in this study. Before presentation, the stimuli in this set were resampled (resample command in MATLAB, The MathWorks, Natick, MA) to place the spectral notch of the 15° stimulus at the BF of the neuron under study. While this resampling sometimes drew the spectral notches outside of the physiological range (6 to 20 kHz,

Musicant *et al.*, 1990; Rice *et al.*, 1992), these stimuli are useful in probing general spectral encoding properties of the neurons. The resampling also changes the stimulus length. Stimuli longer than 400 ms were truncated at 400 ms, while stimuli shorter than 200 ms were repeated to be at least 200 ms long.

The third stimulus set (Fig. 2.1C) was designed to probe the segregation of ABI and ILD cues and disambiguate monaural level responses from true binaural sensitivity. In this set, the frozen noise was again filtered through the spatially averaged HRTF. The ABI of the resultant stimulus was adjusted to range in 5 steps from –8dB to +8dB, referenced to the average system attenuation. This stimulus was then split into two streams that were attenuated relative to one another in such a manner as to preserve the ABI. For this set, ILDs of -16, -8, 0, 8, and 16 dB were used.

The stimuli described above were all created from a single sample of noise, a frozen noise. In this case, the analysis is sensitive to phase-locking to temporal stimulus features, which may be useful information for the auditory system, but may or may not be a useful cue for sound localization. In a number of neurons, the SN/ILD stimulus set was modified to use a different noise on each stimulus presentation. In these cases, each repetition of the 25 SN/ILD stimuli was modified by adding a random vector, sampled from the uniform distribution over $[0,2\pi)$, to the stimulus phases in the Fourier domain. This has the effect of randomizing the temporal structure between sets of stimuli while holding the spectral magnitudes constant. For the analysis done here, these stimuli eliminated information in the stimulus envelope, while preserving information in the spectral magnitudes. To minimize the effects of non-stationarity in the neural response, random waveform sets were presented interleaved with the frozen set.

3. Mutual Information and Bias Estimation

The goal of this work is to quantify the sensitivity of the neural response to various parameters related to sound location, while at the same time making as few assumptions as possible about the encoding process. To do this, techniques from information theory, first introduced by Shannon (1948), are employed. This chapter begins with a general overview of the major ideas from information theory (section 3.1) and discusses how they might be applied to this work (section 3.2). The coding of multiple parameters simultaneously is discussed in section 3.3, and issues of estimation bias are briefly considered in section 3.4. The next sections are devoted to the details of information estimation in various parameters of the neural response, such as spike rate (section 3.5), stimulus decoders (section 3.6), and first spike latency (section 3.7). Finally, the chapter finishes with derivations of the cue-interaction equation (Eqn. 3.13, section 3.8) and the first spike latency information equation (Eqn. 3.31, section 3.9).

The following is only a brief outline of the concepts necessary to apply information theory to the study of neural encoding. A full treatment of the subject can be found in Cover and Thomas (1991) or in MacKay (2002).

3.1 Overview

The whole of information theory can be thought of as stemming from a single concept; namely, that it is possible to quantify the amount of uncertainty in a random variable. This uncertainty is also called the entropy, H, and is defined in the following way. If one considers a (discreet) random variable R that can take on values r from the set {R} with probability p(r), then the entropy of R is given by the equation:

$$H(R) = -\sum_{r\in\{R\}} p(r)\log_2(p(r)). \tag{3.1}$$

The entropy corresponds to our intuition about uncertainty in several respects. First, it is never negative. Second, it is zero only when the probability distribution is degenerate, i.e., collapses to one value with certainty. When all outcomes are equally likely ($p(r)=1/N \; \forall \; r$), the entropy reduces to $\log_2(N)$.

The definition of entropy can be extended to functions of two random variables. For such joint distributions, the definition becomes:

$$H(R,S) = -\sum_{s\in\{S\}}\sum_{r\in\{R\}} p(r,s)\log_2(p(r,s)). \tag{3.2}$$

The major consequence of this definition of entropy for joint distributions is that the uncertainty in two independent random variables is simply the sum of the uncertainty in each individually, i.e., $H(R,S)=H(R)+H(S)$ when $p(R,S)=p(R)p(S)$. In fact, it is this final requirement that constrains the form of the entropy function to a logarithm; the choice of base 2 for the logarithm sets the units of entropy to bits. With this choice of units, the entropy also represents the average number of bits required to convey the value of the random variable over a communication channel.

The conditional entropy, which is the residual uncertainty remaining about random variable R given knowledge of another random variable, S, is defined in a similar fashion:

$$H(R|S) = -\sum_{s\in\{S\}} p(s) \sum_{r\in\{R\}} p(r|s) \log_2 \left(p(r|s)\right), \tag{3.3}$$

where the conditional probability, p(R|S), is defined by the equation p(R|S)=p(R,S)/p(S). With this derivation, it is not hard to derive consistency relationships between the conditional and joint entropies:

$$H(R,S) = H(S) + H(R|S) \tag{3.4}$$

$$H(R,S) = H(R) + H(S|R) \tag{3.5}$$

With the idea of uncertainty placed in quantitative terms, it is now possible to rigorously define information. The mutual information (MI) shared between two random variables is simply the reduction in uncertainty one gains about one variable given knowledge of another:

$$MI(R;S) \equiv H(R) - H(R|S). \tag{3.6}$$

The MI has several properties worth stressing. First, it is symmetric. That is, the information learned about R from S is the same as the information learned about S from R. This can be seen by using the consistency relations of Eqns. 3.4 and 3.5 to yield the following:

$$MI(R;S) = MI(S;R) = H(S) - H(S|R) \tag{3.7}$$

$$MI(R;S) = H(S) + H(R) - H(R,S). \tag{3.8}$$

Second, the MI is bounded, on both sides, such that:

$$0 \le MI(R;S) \le \min\left(H(R), H(S)\right). \tag{3.9}$$

A proof of this result is provided in Cover and Thomas (1991, Chapt. 2).

The mutual information has one further property that will be very important for interpreting the results in this work. It is known as the data processing inequality, and is summarized by the equation

$$\mathrm{MI}(f(R);S) \le \mathrm{MI}(R;S). \tag{3.10}$$

The mutual information between the random variable S and any function of the random variable R must always be less than or equal to the MI between S and R. The equation states in mathematical terms the intuitive result that no amount of post-hoc data processing can increase the information available in the system.

The equations presented here have all been based on the assumption that both of the random variables under consideration, S and R, are discrete. A discussion of the corresponding results for continuous variables is postponed until section 3.7, where the MI of first spike latency codes is considered.

3.2 Applying mutual information to neural encoding

The goal of this work is to understand how sound localization information might be represented by single neurons in the inferior colliculus. For example, one question this thesis will address is whether the type-V units are more sensitive to ITDs than the type-I or type-O units. How can one answer this question, when it is not known how ITDs are represented in the neural response?

The answer, of course, is to use mutual information. Let S be a random variable representing ITD, and R be a random variable representing some feature of the neural

response, such as spike rate or first spike latency. Then one can say that type-V units are more sensitive to ITDs than type-O units if the average MI carried by type-V units about ITD is significantly more than the average MI carried by type-O units.

The advantage of using MI as a measure of neural sensitivity is that it allows one to make very few assumptions about the form of the neural representation. For example, if the response variable is spike rate, it does not matter if changes in ITD lead to decreases or increases in the firing rate, as both sorts of changes are captured by the MI definition. In fact, the mean firing rate need not change at all; if ITD manipulations lead to changes in the variance of the firing rate without affecting the mean firing rate, MI would capture this sensitivity.

The caveat is that a response variable must still be chosen, and its choice still represents an assumption about neural coding. In theory, one could use the entire spike train, characterized as a set of spike times, as the response variable. In practice, however, one could record one hundred or even one thousand spike trains from the same stimulus and not see the same spike train twice. The multi-dimensional spike train space is too large to probe experimentally.

Therefore, it is necessary to come up with low-dimensional descriptors of the spike train space that retain the relevant features relating to stimulus discrimination while still being of low enough dimension to probe experimentally. This is an open question in neuroscience research: what features of spike trains are used to convey information? This study attempts to answer this question in a limited and well-defined setting. By quantifying the MI between sound localization cues and several low-dimensional descriptors of spike trains, such as firing rate and first spike latency, this study attempts

to gain some understanding of the neural code for sound localization in the inferior colliculus.

3.3 Mutual information in two variables

3.3.1 Mutual information in two stimulus variables

The stimuli used in this study are all characterized by two sound localization cues. Thus, the equations above for MI need to be extended. Expanding Eqn. 3.6 by inserting the definitions of the entropies yields

$$\mathrm{MI}(S;R) = \sum_{s\in\{S\}}\sum_{r\in\{R\}} p(s,r)\log_2\left(\frac{p(s,r)}{p(s)p(r)}\right). \tag{3.11}$$

If X and Y represent the two localization cues, then each stimulus is characterized by a vector of parameters, **S**=(X,Y). The information between the response and both localization cues, MI(**S**;R), becomes MI(X,Y;R), and the sum over **S** becomes a sum over X and Y:

$$\mathrm{MI}(X,Y;R) = \sum_{x\in\{X\}}\sum_{y\in\{Y\}}\sum_{r\in\{R\}} p(x,y,r)\log_2\left(\frac{p(x,y,r)}{p(x,y)p(r)}\right). \tag{3.12}$$

Note that the MI is still being considered between just two random variables, R and the vector **S**, so the information is still constrained to be positive. For notational convenience, the MI between R and both localization cues X and Y will be referred to as $\mathrm{MI}_{\mathrm{FULL}}$, and the MI between R and one localization cue will be known as MI_X or MI_Y. Eqns. 3.11 and 3.12 for $\mathrm{MI}_{\mathrm{FULL}}$ will be used interchangeably in the following text.

The relationship between the individual cue information and the full stimulus information is important in understanding how individual neurons represent multiple cues. A full derivation, provided in section 3.8, yields the cue-interaction equation:

$$MI_{FULL} = MI_X + MI_Y + MI(X;Y|R), \tag{3.13}$$

which holds when MI(X;Y)=0, as is the case for these stimuli. The mutual information between the response and the full stimulus set is not necessarily equal to the sum of the information about the individual cues. Rather, since MI(X;Y|R) must be non-negative, MI_{FULL} must be greater than or equal to the sum of the individual cue information.

This is perhaps a confusing result for those familiar with papers on population information (Petersen *et al.*, 2001; Reich *et al.*, 2001a; Puchalla *et al.*, 2005). In that literature, it is often noted that the MI provided by two neurons about a stimulus set can be less than or equal to the sum of the MI provided by each neuron individually, as would happen if there were redundancy in the neural response. These population studies should not be confused with the current case, which is concerned with the representation of multiple cues within single neurons, only.

The difference between MI_{FULL} and the sum of the individual cue information, MI(X;Y|R), is known as the confounded information (introduced by Reich *et al.*, 2001b), and is related to the (lack of) independence in the neural response. For example, a non-zero value of the confounded information will occur when the spike count in response to parameter X changes depending on the value of parameter Y. Though the stimulus parameters X and Y are originally independent of one another, conditioning on the spike count can induce mutual information between them. From a coding point of view, the information a neuron provides about X can be increased by independent knowledge of Y,

even though the stimuli are formally independent *a priori*. More importantly, non-zero confounded information means that the cues cannot be decoded independently.

The percentage of the information that is devoted to encoding a particular localization cue is known as the relative information, and serves as the basis for analyzing the coding interaction of multiple cues. It is computed as $MI_X\%=MI_X/MI_{FULL}$ (%). Rewriting the cue-interaction equation in terms of relative information yields:

$$100 = MI_X\% + MI_Y\% + CI, \tag{3.14}$$

where $CI=MI(X;Y|R)/MI_{FULL}$ (%) is the percentage of confounded information. By comparing the relative informations for different codes, one can test hypotheses about the response variables used to encode particular cues.

3.3.2 Mutual information in two response variables

The preceding section dealt with the issue of two stimulus variables. This section is quite similar, but deals with the issue of two response variables. It may be profitable to look at the information encoded about the localization cues in a joint code consisting of two response variables. For example, these response variables might be rate and first spike latency, and the question could be how much information they jointly contain about the stimulus set.

Following the argument of the last section, the response can now be considered a vector, such that $\mathbf{R}=(R_1,R_2)$. The information contained about the stimuli in this joint code, MI(R1,R2;S), will be referred to as MI_{JOINT}. The derivation provided in section 3.8 for the cue-interaction equation follows exactly for looking at the information in a joint

code, with the exception that we cannot assume that the coding variables are independent as we could with the stimulus variables. Thus, $MI(R_1;R_2)$ is not, in general, zero, and the interaction equation becomes, with some rearrangement:

$$MI_{JOINT} - MI(R_1;S) - MI(R_2;S) = MI(R_1;R_2|S) - MI(R_1;R_2) \quad (3.15)$$

The left hand side of this equation is referred to as the Redundancy/Synergy interaction, following the discussion of Chechik and colleagues (2001).

Essentially, the difference between the information provided by a joint code and the sum of the information provided by each code alone can be positive or negative. When negative, the individual codes are said to be redundant, and when positive, the codes are synergistic. Redundancy is caused by a lack of independence in the chosen coding variables, i.e., when $MI(R_1;R_2)$ is not zero. Synergy, on the other hand, is caused by information between the response variables *induced* by the stimuli, $MI(R_1;R_2|S)$. In analogy with the confounded information, synergy occurs when neither R_1 nor R_2 on their own uniquely identify the stimulus, but their combination does.

3.4 Bounds and bias of mutual information estimates

In Eqn. 3.9, it was noted that the MI between two random variables must be non-negative and less than or equal to the maximum of the two entropies. As discussed in Chapt. 2, all of the stimulus sets considered in this work are composed of 25 stimuli presented an equal number of times, so MI_{FULL} is bound from above by $\log_2(25) \cong 4.64$ bits. Similarly, since there are five equally likely values of each of the individual localization cues, MI_X and MI_Y are bound from above by $\log_2(5) \cong 2.32$ bits.

Because probability distributions are hard to characterize, estimates of mutual information are notoriously noisy (Macrae, 1971; Golomb *et al.*, 1997; Paninski, 2003). The noise can be divided into two components: variance and bias. It is not difficult to understand the origins of the variance component. Sampling twice from the same underlying probability distribution will tend to yield different outcomes; these different outcomes translate into different estimates of the MI. Different estimators of MI will be more or less sensitive to sampling noise of this type.

The origin of the bias in MI estimates is less straightforward. Bias is defined as the difference between the expected value of an estimator and its true value, as summarized by the equation

$$\text{Bias} \equiv \mathrm{E}\left[\mathrm{MI}_{\text{est}}\right] - \mathrm{MI}_{\text{true}}. \tag{3.16}$$

While it is often said that MI estimates are positively biased because the MI cannot be negative (and therefore noise will lead to an overestimation of MI, on average), this argument does not hold in general, as the MI is bounded in both directions. Rather, the bias stems from the particular form of the logarithmic non-linearity in the MI equations.

Consistent estimators of MI will converge to the true value of the information as the number of data repetitions tends to infinity. Good estimators of MI will converge quickly. For any study using information theory, it is necessary to understand the approximate size of the bias and the variance, and their behavior in terms of the amount of data collected. As these values depend on how the MI is estimated; a full discussion of these issues is given in the next few sections describing different MI estimation techniques.

3.5 Mutual information in spike rate

Spike rate is probably the most obvious neural feature that changes in response to stimulus presentation. This section will start by introducing the most commonly used estimator of spike rate MI, the maximum likelihood estimate, in section 3.5.1. The most popular debiasing techniques will be introduced and compared in section 3.5.2. Finally, section 3.5.3 will introduce a new estimation technique, the fill-in method, that is shown to yield more accurate estimates of spike rate MI in certain situations.

3.5.1 Rate-based MI and the maximum likelihood estimate

Estimating the joint probability distribution for a spike count code (or a spike rate code; given a fixed analysis window the two are interchangeable) is fairly straightforward: if $N_{obs}(s,r)$ represents the number of times spike count *r* was observed during presentation of stimulus *s* and $N_{reps}(s)$ represents the number of presentations of stimulus *s*, then the conditional probability p(R=r|S=s) is given by

$$p(R=r|S=s)=\frac{N_{obs}(s,r)}{N_{reps}(s)}. \tag{3.17}$$

All of the data in this study were taken with equal numbers of stimulus presentations (i.e., $N_{reps}(s)=N_{reps}$, and therefore $p(s)=1/N_{reps}$), so for the rest of this work N_{reps} will refer to the number of repetitions of any given stimulus. From the definition of conditional probability, the joint probability distribution is given by

$$p(R=r,S=s)=p(R=r|S=s)p(s)=\frac{N_{obs}(s,r)}{N_{reps}^2}. \tag{3.18}$$

This is known as the maximum likelihood estimate (MLE) of the probability, since this estimate of the probability is the one most likely to have yielded the observed spike count (see Stark and Woods, 2002, Chapt. 5 for review); plugging this estimator into the definition of MI_{FULL} in Eqn. 3.11 yields the MLE of the spike count MI. The marginal probability distribution of the response, p(R), is derived from the joint probability distribution by

$$p(R) = \sum_{s=1}^{25} p(R, S = s), \tag{3.19}$$

where *s* indexes each one of the 25 stimuli in the full stimulus set.

The joint probability for computing the individual cue information, MI_X, is also derived from Eqn. 3.18. The stimulus index *s* is broken into its component parts *x* and *y*, which each range over the 5 values of one localization cue. Then p(X,R) is calculated as

$$p(X, R) = \sum_{y=1}^{5} p(R, X, Y = y), \tag{3.20}$$

and similarly for p(Y,R). Note that with these estimates of p(X,R) and p(Y,R), Eqn. 3.13 holds exactly (see section 3.8).

3.5.2 Debiasing schemes for rate-based MI estimates

The MLE estimate of the spike count MI is typically biased (Treves and Panzeri, 1995); in fact, the bias often contributes more to the estimation error than the variance. This bias can be demonstrated in simulations, where the spike counts for each stimulus *s* are assumed to come from a Poisson distribution with mean $\lambda(s)$. While real spike count distributions are not exactly Poisson (since the Poisson distribution fails to account for

the refractory period), the Poisson distribution represents a simple model for spike count statistics and is a fairly close match to spike count distributions in the auditory nerve at low spike rates (Weiss, 1964; Moore *et al.*, 1966).

The simulation setup is as follows. N_{reps} sample spike counts are drawn for each of 25 stimuli from a Poisson distribution with mean $\lambda(s)$. These counts are used to derive the MLE of the MI, as detailed in Eqns. 3.11 and 3.16 – 3.18. The model experiment is run 50 times, with a different set of spike counts drawn each time, and the results are compared to the real MI, calculated numerically using the Poisson distribution for the conditional probabilities. Figure 3.1A shows the results of one such simulation plotted as a function of N_{reps}, where the real MI (black, dotted line) is ~ 1.2 bits (see figure caption for lambda values). The blue line represents the mean (± 1 SD) of the 50 runs. The difference between the mean of the simulated runs and the real MI is the bias of the MLE in this simulation. This bias is not only a function of the number of stimulus repetitions, but is also a function of the real MI, as shown in Fig. 3.1B.

Two things are immediately clear from Fig. 3.1A. First, the MLE of the spike count MI is always positively biased; second, this bias is much larger than the error bars. For the simulation of Fig. 3.1A, approximately 60 repetitions of every stimulus are required to yield MI measurements within 20% of the real value.

There are many studies in the literature devoted to eliminating the bias in these kinds of MI estimates. The three most commonly accepted are the bootstrap procedure (Efron and Tibshirani, 1998), the analytical correction (Panzeri and Treves, 1996), and the extrapolation method (Strong *et al.*, 1998). The first and last of these are data driven, in that they estimate bias directly from the measured set of spike counts. The second

method is a function of the number of stimuli, number of stimulus repetitions, and the number of different responses only, and so depends little on the actual data. Although other methods for bias estimation have been proposed (Optican *et al.*, 1991; Paninsky 2003), they have yet to receive much use by experimenters and won't be expounded upon here.

To estimate bias with the bootstrap procedure (presented in detail in Efron and Tibshirani, 1998), one draws a sample set of spike counts for each stimulus by sampling (with replacement) from the set of observed spike counts for that stimulus. The sample set should have as many observations as the original set; the relative frequencies of the counts will be different, however, because of the replacement. These bootstrapped spike counts are used to produce a bootstrap estimate of the MI. This procedure is repeated multiple times (typically, hundreds), until the mean of the bootstrapped MI estimates converges. The bias is then estimated as the difference between the actual MI estimated from the original stimuli and the mean of the bootstrapped MI estimates,

$$\mathrm{Bias}_{\mathrm{boot}} = \mathrm{MI}_{\mathrm{data}} - \left\langle \mathrm{MI}_{\mathrm{boot}} \right\rangle . \tag{3.21}$$

Figure 3.2A shows a plot of the simulated MI estimates, after debiasing with the bootstrap method (using 500 bootstrap repetitions), as a function of N_{reps}. These estimates are clearly superior to their undebiased counterparts, in that only about 25 repetitions are required to converge to within 20% of the real MI. Even so, there is clearly residual bias that is unaccounted for by the bootstrap method. This residual bias is plotted as a function of N_{reps} for simulations with different real MI values in Fig. 3.2B.

The analytical method, presented in detail by Panzeri and Treves (1996), derives a first order approximation to the bias from first principles. Essentially, the bias is expanded as a power series of $1/N_{reps}$,

$$\text{Bias}_{\text{Anal}} = \sum_{i=1}^{\infty} \frac{C_i}{N_{reps}^i}. \tag{3.22}$$

The first term in this series is found not to depend strongly on the data, such that

$$\text{Bias}_{\text{Anal}} = \frac{1}{2\ln(2) N_{stim} N_{reps}} \left\{ \sum_{s \in \{S\}} |R_s| - |R| - (N_{stim} - 1) \right\} + O\left(\frac{1}{N_{reps}^2} \right), \tag{3.23}$$

where $|R_s|$ represents the number of different responses to stimulus s, $|R|$ represents the total number of response categories, N_{stim} represents the number of different stimuli (25 for these experiments), and N_{reps} represents the number of presentations of each stimulus. Note that this equation assumes that each stimulus is presented an equal number of times, as is true for this work.

Figure 3.2A shows a plot of the simulated MI estimates, debiased using Eqn. 3.23, as a function of N_{reps} (solid green line). This debiasing technique works the best in terms of absolute convergence, where only between 10 and 15 repetitions are required for the estimated MI to convergence to within 20% of the real MI. The debiased estimates also show some residual bias (Fig. 3.2C), but in this case, the residual bias can be either positive or negative, depending on the number of stimulus repetitions and the real MI. This method does the best when the real MI is smallest, and gets progressively worse as the real MI increases.

The extrapolation method of debiasing has its roots in the analytic expansion represented in Eqn. 3.22 (Strong *et al.*, 1998). The data is divided into multiple fractions of various sizes (1 part, 2 parts, 3 parts, etc.), and the MI is estimated for each data

fraction. Plotting the estimated MI as a function of the data fraction yields a curve that can be fit with a linear (or at most, quadratic) function. The y-intercept of this curve, representing the MI extrapolated to $1/N_{reps} \rightarrow 0$, or $N_{reps} \rightarrow \infty$, is the MI estimate with infinite data, which has no sampling bias. In practice, care must be taken not to allow the data fractions to become too small, or the linear/quadratic fits will be poor and the variance of the MI estimate will be large.

The solid red line of Fig. 3.2A shows a plot of the simulated MI estimates debiased with this extrapolation procedure. The results are substantially like the bootstrapped-debiased results, including the magnitude of the residual, uncorrected bias plotted in Fig. 3.2D.

The standard deviations of the MI estimates made using each debiasing method are compared for several different values of the real MI in Fig. 3.3. Of course, the standard deviation of each estimate is a function of the amount of data collected; the standard deviation values are compared for N_{reps}=16 (Fig. 3.3A), N_{reps}=64 (Fig. 3.3B) and N_{reps}=256 (Fig. 3.3C). The bootstrap method tends to have a slightly smaller variance than the other debiasing techniques.

Ultimately, the experimental concern is not with either the bias or the variance in MI estimates; it is with the error. The average mean squared error that one measures in the simulation is

$$\text{Error} = \text{Bias}^2 + \text{Variance}\,. \tag{3.24}$$

Figure 3.4 shows the average error in the simulations plotted as a function of N_{reps} for all of the estimation methods described so far. The error has been averaged over the four simulation runs with different information contents, as described in Fig. 3.1. The

analytical debiasing method tends to outperform the other methods, although the difference becomes quite small after 50 repetitions or so.

3.5.3 The fill-in method: a new debiasing technique

The debiasing techniques presented above all work as corrections to estimates of spike count MI made with the MLE. This section details a new estimation technique that I call the fill-in method. It differs from the above techniques by not using the MLE as its base; rather, the probability distributions are estimated in a different way so as to produce a smaller bias in the spike count MI from the start. Debiased versions of the fill-in method will be shown to be more accurate than current estimation techniques for small sample sizes.

In the MLE of the spike count MI, unobserved spike counts are given zero probability (essentially, they are deemed impossible). These zero probability events contribute greatly to the MI, because observing that number of spike counts necessarily rules out certain stimuli. However, under limited sampling constraints it is possible that there simply were not enough occasions to observe each possibility, and therefore these events may not be impossible, merely less probable. Of course, given the data available, it is impossible to distinguish between these two possibilities. The fill-in method differs from the MLE method by having a more stringent requirement for assigning a zero probability spike count.

The main assumption of the fill-in method is that, for any given stimulus, spike counts with values between the minimum and maximum observed cannot occur with zero

probability. As an example, if for a certain stimulus one observes 5 spikes on one presentation and 7 spikes on another, it is assumed that getting 6 spikes is not impossible. Note that this does not rule out the possibility of a bimodal spike count distribution, it just means that the probability cannot go to zero between the modes.

The conditional probability of observing spike count *r* given stimulus *s* for the fill-in estimate of MI is

$$p(R = r|S = s) = \frac{N_{obs}(s,r) + \frac{1}{2}}{Z}, \tag{3.25}$$

where $N_{obs}(s,r)$ is the number of times spike count *r* was observed while presenting stimulus *s* and *r* ranges from the maximum observed spike count to the minimum observed spike count. Z is a normalization constant that ensures p(R|S) sums to unity, and can be calculated as

$$Z = N_{reps} + \frac{(\max(r) - \min(r) + 1)}{2}. \tag{3.26}$$

As an illustrative example, say that in four repetitions of a stimulus, one observes spike counts of 3, 5, 7, and 7. Given that the maximum observed count is 7, the minimum is 3, and N_{reps} is 4, Z = 6.5. Under the fill in method, all spike counts between 3 and 7 would be possible, with values:

r	$(N_{obs} + 0.5)/Z$	p(r)
3	$1.5/6.5$	0.23
4	$0.5/6.5$	0.08
5	$1.5/6.5$	0.23
6	$0.5/6.5$	0.08
7	$2.5/6.5$	0.38

(3.27)

Figure 3.5A shows a comparison of the MLE and fill-in methods of MI estimation, using the same simulation parameters as in Fig. 3.1A. Clearly the fill-in method

outperforms the MLE method, especially at low number of stimulus repetitions. The fill-in estimates converge to within 20% of the real MI for 25 stimulus repetitions, compared to ~60 repetitions for the undebiased MLE. Repeating these simulations for several values of the real MI (Fig. 3.5B) shows that the fill-in estimates of MI are usually positively biased, although negative biases are observed for large N_{reps} and real MI. Comparison with Fig. 3.1B shows that the bias in the fill-in method is less dependent on the real MI than the MLE method is.

Both the bootstrap and the extrapolation method can be used to debias fill-in estimates of MI. (Analytical estimates of bias for the fill-in method are complicated by the change in the number of possible responses for each stimulus.) Figure 3.6A shows how fill-in estimates of MI converge with sample size when debiased with either the bootstrap or extrapolation techniques. Bootstrap debiased MI values from the MLE method are shown for comparison. For low numbers of stimulus repetitions, the debiased fill-in method outperforms the debiased MLE method. Bootstrap debiased fill-in estimates tend to underestimate the real MI for moderate numbers of stimulus repetitions. This is shown in more detail in Fig. 3.6B, where the residual bias in fill-in MI estimates debiased with bootstrap is shown for several values of the real MI. The amount of negative residual bias is a function of the information content of the simulation. The residual bias in the extrapolation-debiased estimates (Fig. 3.6C) is less dependent on the real MI, but is also larger than the bootstrap values.

The average error in fill-in MI estimates is shown as a function of sample size in Fig. 3.7. The best case debiasing method based on the MLE is shown for reference. For small numbers of stimulus repetitions, the fill-in estimates far outperform the best case

based on the MLE. (See Fig. 3.4 for comparison, and note the scale differences.) In fact, bootstrap-debiased fill-in estimates produce almost an order of magnitude less error than the analytical-debiased MLE values for the smallest number of repetitions tested. At more moderate sample sizes (~60 repetitions), the average errors in all of the estimates are comparable.

3.6 Mutual information in a stimulus decoder

A stimulus decoder takes as its input a spike train, and as its output returns an estimate of the stimulus parameters that likely produced the input spike train. This estimate can be something simple, such as returning the stimulus with the closest mean firing rate to the input spike train, or it can be some arbitrarily complex function of the input spike times. Stimulus decoders based on artificial neural networks (Middlebrooks *et al.*, 1994), Bayesian estimation (Wiener and Richmond, 2003), and distance metrics (Victor and Purpura, 1997) have been used to investigate neural coding.

This section details how the MI of a stimulus decoder is estimated. Decoder-based estimates of MI are popular because while the stimulus estimate itself can be a complicated function of the set of spike times, the description of the response is low-dimensional, with the number of possible values corresponding to the number of stimuli in the set. Thus, decoder-based methods make it possible to estimate the MI in spike timing without paying the cost of huge amounts of data. Of course, there is a downside: any MI estimate based on a decoder is subject to the data processing inequality of Eqn.

3.10, and must be considered as a lower bound to the actual information present in the spike trains.

3.6.1 Estimating MI from a hard partition decoder

For the purposes of this discussion, the details of the decoding process do not matter. Under the assumption of a hard decision process, the decoder will take a spike train and output one stimulus estimate. (The fuzzy decision process, where the decoder outputs multiple stimulus estimates with different probabilities, will not be considered.) Since the stimuli are themselves composed of two localization cues, each stimulus estimate is an estimate of each localization cue. For simplicity, in the following discussion the stimuli will be indexed by *s* and the stimulus estimates by *r*, where *s* and *r* run over all 25 localization cue pairs.

Once the decoder is run on every spike train, the results can be gathered together into a confusion matrix, **N**, where each element N_{sr} is the number of times that a spike train recorded during presentation of stimulus *s* was estimated to have come from stimulus *r*. To compute the MI from the confusion matrix, let the responses, R, be the estimated stimuli, and the stimuli, S, be the actual stimuli. The conditional probability is found by normalizing by the number of stimuli N_{reps}

$$p(R = r|S = s) = \frac{N_{rs}}{N_{reps}}, \tag{3.28}$$

and the joint probability is found by multiplying Eqn. 3.27 by p(s)

$$p(R = r, S = s) = \frac{N_{rs}}{N_{reps}^2}. \tag{3.29}$$

Note that these equations are the same as the ones used for the MLE of the spike count MI (Eqns. 3.16 and 3.17), and represent the MLE of the decoder MI.

The marginal probability distributions p(R) and p(X,R) are derived from the joint distributions in exactly the same manner as for the rate case, using Eqns. 3.18 and 3.19. This leads to a subtlety in the confusion matrix analysis for individual localization cues. An individual localization cue can take on only one of five possible values, whereas the estimated stimulus still runs over all 25 possible pairs of localization cue values. Thus, $p(x,\mathbf{r}=[r_x,r_y])$ represents the probability that a spike train from one of the stimuli with localization cue X=x was estimated to have been come from a particular stimulus with localization cue $X=r_x$ and localization cue $Y=r_y$. If one is interested only in the information carried about localization cue X, why not group all the Y responses into one bin? The reason is that then the cue-interaction equation (Eqn. 3.13) would not hold. This is discussed in greater detail in section 3.8.

3.6.2 Debiasing confusion matrices

There are two possible sources of bias in decoder estimates of MI. The first lies in the decoding process itself; depending on the details of the decoder, it is possible that the assignment of spike trains to stimulus estimates results in a biased distribution of responses. The second source of bias is in the estimation of MI from the confusion matrix. While it is ideal to compensate for both sources of bias, in practice it is sometimes difficult to estimate bias in the decoding process, as some decoding algorithms are too computationally intensive to run multiple times. This section is

concerned mainly with the bias in the second step, when the MI is estimated from the confusion matrix, although the effects of bias in the decoding process will be modelled.

Here, the decoding process is looked upon as a black box that is characterized by a confusion matrix with a set of unknown probabilities. The actual confusion matrix output with the available data represents the experimental sample, and it is this sample that is manipulated to form bias estimates. There are two popular ways of performing this type of debiasing. One is through a bootstrap process (Efron and Tibshirani, 1998) and the other is through random reassignment (Victor and Purpura, 1997).

The bootstrapping process for confusion matrices is similar to that for the rate case. Each stimulus is handled independently, by drawing N_{reps} samples (with replacement) from the estimated responses corresponding to that stimulus. These responses are placed into a new, bootstrapped confusion matrix, resulting in a different distribution of response probabilities while holding the stimulus probabilities constant. A bootstrapped estimate of the MI is made from the bootstrapped confusion matrix using Eqn. 3.29. This process is then repeated until the mean of the bootstrapped MI estimates converges to a stable value, and the bias is estimated using Eqn. 3.21.

The random reassignment method addresses the bias in a more analytical way, by measuring the amount of MI in a confusion matrix with no relationship between stimuli and responses. To do this, the N_{reps} responses for each stimulus are assigned to the response categories for that stimulus at random, and the MI is estimated from the randomized matrix. This process is continued until the mean of the random reassignment MI values converges; this mean is an estimate of the bias.

Figure 3.8 shows the random reassignment estimate of the bias as a function of N_{reps} for the 25x25 confusion matrices used in this study. Since the bias estimate does not depend at all on the data (it only depends on N_{reps} and the size of the confusion matrix), this curve serves as the bias estimate regardless of the real underlying probability distribution. The real bias, however, is a function of the underlying probability distribution. This method corrects for bias best when the real underlying probability distribution is close to uniform.

To test the accuracy of the MI estimates for confusion matrices, a "real" decoder matrix is created as a set of probabilities, p(r|s), determining the likelihood that stimulus *s* is estimated as stimulus *r*. A confusion matrix is created from this decoder matrix by assigning N_{reps} samples for each stimulus to the estimate categories with the specified probabilities. This can be done by dividing the real line from 0 to 1 into 25 sections of length p(r|s). Each sample is placed into a response category according to the section that a random number drawn from the uniform distribution falls into.

The decoder in these simulations is specified by the set of equations

$$p(r|s) = \begin{cases} \alpha, & r = s \\ \dfrac{1-\alpha-\beta}{24}, & r \neq s \\ +\beta, & r = 1 \end{cases} \tag{3.30}$$

where α and β are parameters that can be adjusted to mimic different decoder accuracies and failure modes. The plus sign on the last condition, r=1, means that the value of p(r|s) specified by the previous conditions is incremented by β. In other words, if s=1, p(1|s)= $\alpha+\beta$, otherwise p(1|s) is equal to $(1-\alpha-\beta)/24+\beta$. This always ensures that the elements of

p(r|s) sum to 1. Increasing α serves to increase the decoder accuracy, and therefore increase the information content. When β is zero, the decoder fails in such a way that incorrect stimuli are assigned to other bins randomly. When β is non-zero, the decoder overweights the r=1 response, representing bias in the decoding process. Actual decoders tend to suffer from a mixture of these two types of failures.

Figure 3.9A, C, and E shows several examples of information estimates for confusion matrices simulated using Eqn. 3.30 and debiased by both bootstrap and random reassignment. The simulations differed in their α values, which determine the amount of real information; β is zero in all of these cases. Bias correction with the random renormalization scheme outperforms bootstrap bias correction estimates, as can be seen in the error plots of Fig. 3.9B, D, and F. However, while bootstrap tends to underestimate bias, random renormalization tends to overestimate it. Random renormalization perfectly accounts for the bias in Fig. 3.9A, because with α=0.04 and β=0, the probability values of the confusion matrix are all equal. Hence, stimuli are assigned randomly to responses, which is exactly the assumption behind the random renormalization procedure.

Figure 3.10 shows a similar set of plots, but in this case β has been set to 0.5, so the decoder tends to overweight its estimate of stimulus 1. As the information in the system increases, the random renormalization method produces information estimates that get progressively worse. The bootstrap bias estimates adapt to changes in the underlying probability distribution better than the random renormalization bias estimates.

3.7 Mutual information in first spike latency

Up to this point, the allowed response variables could only take on discrete values. Spike arrival times are inherently continuous, however, and direct estimation of the MI in spike arrival times will have to take this into account. In particular, the first spike latency has been shown to carry stimulus information in several sensory systems (Furukawa and Middlebrooks, 2002; Panzeri *et al.*, 2001; Reich *et al.*, 2001c). This section considers the different aspects of first spike latency information (section 3.7.1) and details two methods of estimating the MI in the timing of the first spike latency: the kernel density estimate (KDE) method (section 3.7.2), and the binless method (section 3.7.3). Decoder-based methods can also be used to get a lower bound on first spike latency information; the results of the previous section can be used to derive the MI from those methods.

3.7.1 The breakdown of first spike latency information

The first spike latency can actually carry information in two independent mechanisms that must be distinguished for proper estimation. First, not all stimuli cause the neuron to spike. Thus, the first question one must consider when investigating the information in the first spike latency is whether or not there was actually a spike. The answer to this question is characterized by the probability, $p_{spike}(s)$, that each stimulus elicited at least one spike. Since $p_{spike}(s)$ can differ from stimulus to stimulus, it can carry information; this information will be referred to as MI_{count}. Second, given that the stimulus did cause at least one spike, the probability distribution of this spike arrival time can vary from stimulus to stimulus and so can also carry information. This information will be referred

to as MI_{timing}. The breakdown of first spike latency information is represented by the equation

$$MI_{FSL} = MI_{count} + p_{count} MI_{timing}, \tag{3.31}$$

where p_{count} is the mean value of $p_{spike}(s)$ over all stimuli, and MI_{FSL} represents the total information between first spike latency and the stimuli. A derivation of this equation is provided in section 3.9.

MI_{count} is calculated in a very straightforward manner. The response random variable is a binary variable representing the presence or absence of a spike, and hence is discrete. It can be estimated using a maximum likelihood approach, where $p(r=1|s)=N_{sp}(s)/N_{reps}$, $p(r=0|s)=1-p(r=1|s)$, and $N_{sp}(s)$ is the number of times stimulus *s* elicited a spike. MI_{count} can be debiased using any of the standard rate debiasing methods (bootstrap, analytical, or extrapolation), but in practice the bias is usually quite low, because there are only two response categories. Assuming a worst-case scenario where every stimulus yields at least one instance of each response category, for $N_{reps}=70$ (the median number of stimulus presentations in this work), the analytical method yields a bias estimate of approximately 0.01 bits.

MI_{timing} is a bit more difficult to estimate, because while the stimulus random variable is still discrete, now the response random variable is a time, which is continuous. Mutual informations involving one discrete and one continuous random variable can be calculated with the equation

$$MI(R;S) = \sum_{s \in \{S\}} p(s) \int_R f(r|s) \log_2 \left(\frac{f(r|s)}{f(r)} \right) dr, \tag{3.32}$$

where f(r|s) represents the conditional probability *density* of r given s. The marginal probability density of r, f(r), is calculated as

$$f(r) = \sum_{s \in \{S\}} p(s) f(r|s). \tag{3.33}$$

The problem of computing information in spike timing, then, boils down to the problem of estimating the probability density functions. The obvious way of doing this is via a histogram method, where the time axis is binned and the number of occurrences of the spike arrival times in each bin determines the density function. The histogram method presents some difficulties, however, that are expounded upon in the next two sections. In section 3.7.2, an alternative to the histogram-based method of estimating the probability densities is considered, and in section 3.7.3, a method is presented that ultimately allows MI_{timing} to be calculated from the data directly, without resorting to a numerical estimate of the integral in Eqn. 3.32.

3.7.2 Kernel density estimates of MI

The Kernel density estimate (KDE) of MI derives its name from a general method of estimating continuous probability densities from a set of discrete samples (Parzen, 1962). Briefly, the conditional probability density of the timing of the first spike latency is estimated by taking every observed first spike time and convolving it with a kernel K(r), such that

$$f(r|s) = \frac{1}{n(s)} \sum_{i=1}^{n(s)} K(r - t_i). \tag{3.34}$$

Here, n(s) is the number of first spike times observed while presenting stimulus *s*, and t_i represents the i^{th} observed spike time. For Eqn. 3.34 to be valid, the kernel K(r) must be normalized such that its integral is 1. In practice, these kernel-based methods have been found to produce estimates of probability density functions that converge faster than traditional histogram-based techniques (Glavinović, 1996).

While KDEs can be developed using any properly normalized kernel, in this work only the gaussian kernel is considered, as it has only one parameter and there is no *a priori* reason to use something more complicated. The kernel is specified by

$$K(r) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-r^2}{2\sigma^2}\right)}, \tag{3.35}$$

where σ represents the standard deviation of the kernel. Care must be used in choosing σ. Kernel widths that are too wide will result in probability functions that are too smooth, leading to underestimation of the timing information. Conversely, kernel widths that are too small will lead to probability distributions that are too bumpy, causing overestimates of the timing information.

One can gain an understanding of the behavior of the MI estimate as a function of the Kernel width by simulating a system where the MI is known. Assume that presentation of stimulus *s* always leads to a spike whose arrival time is completely specified by a gaussian distribution with mean αs and standard deviation 1. That is, p(r|s)=Normal(αs,1), and the only effect of the stimulus is to shift the average arrival time of the spike via the mean of the gaussian. α is a scale factor that controls the separation of the distribution means, and therefore controls the amount of information in the system.

For 25 stimuli, indexed by *s*=[1,25] with α=1, the real MI can be shown numerically (using Eqn. 3.32) to be ~2.1 bits.

Each run of the simulation involves drawing N_{reps} samples from each of the gaussian distributions and estimating the MI using Eqns. 3.31-3.34 with a fixed kernel width. The simulation is run 50 times, to measure the mean and standard deviation of the MI estimates. Figure 3.11 shows the results of these simulations for several different kernel widths and sample sizes, with α=1. If the kernel width is incorrectly chosen, the MI can be overestimated or underestimated by large amounts, regardless of the amount of data collected.

Fortunately, a method exists for computing the optimal kernel width from the collected data (Schraudolph, 2004). Essentially, the sample points observed in response to one stimulus are divided into two parts, U and V. The first subsample, U, is used to build an estimate f'_U of the probability density function using Eqn. 3.11. The (log) likelihood of observing subsample V given the density f'_U is calculated as

$$L = \ln\left(\prod_{r_i \in V} f'_U(r_i)\right) = \sum_{r_i \in V} \ln\left(f'_U(r_i)\right)$$

$$= \sum_{r_i \in V} \ln\left(\sum_{r_j \in U} K(r_i - r_j)\right) - |V| \ln |U|, \qquad (3.36)$$

where |U| and |V| represent the number of sample points in subsamples U and V, respectively. The optimal kernel width, σ_{opt}, is defined as the kernel width that maximizes L. In other words, the kernel is chosen such that the estimated density from subsample U maximizes the probability of having observed subsample V.

In practice, the subsamples are chosen through the leave-one-out method, where V contains one sample and U contains the others. Likelihood functions are computed for each different subsample V, and σ_{opt} is chosen as the kernel width that maximizes the average likelihood function. This procedure is then repeated for every stimulus, resulting in a σ_{opt} for each.

Figure 3.12 shows examples of how the MI estimates using the optimal kernel width converge to the real MI value as a function of N_{reps}. While it appears that the errors converge quite slowly with increasing N_{reps} (Fig. 3.12D), note the absolute value of the error is actually much lower than in previous figures (e.g., Figs. 3.4 and 3.9). The average optimal kernel width decreases as N_{reps} increases (Fig. 3.13), indicating that as more data points are collected, less averaging is needed. Each point becomes a much more local estimate of the probability density function.

3.7.3 Binless estimates of MI

The binless estimate of MI is interesting in that it allows the information between a discrete and a continuous random variable to be estimated without resorting to the calculation of probability density functions. Developed by Victor (2002), it was originally formulated as a way to get around the histogram-based method. The problem with the histogram method is in determining how the time axis should be partitioned. Bin widths that are too wide tend to underestimate the information by smearing high-resolution responses, while bin widths that are too small tend to overestimate the information due to sample size bias problems. While methods exist in the literature to

estimate the "optimal" bin size (see, for example, Nelken *et al.*, 2005), ideally one would prefer a method that is less dependent on sample size.

While a full derivation of the binless method is beyond the scope of this work, the major steps of the derivation will be discussed here in a qualitative manner. First, the entropy of the first spike times can be estimated by approximating the continuous probability density as a series of delta functions located where spikes are observed. Each delta function is of size 1/N, where N is the number of observed spikes, so the entropy of the first spike density becomes

$$H(R) \approx -\sum_{j=1}^{N} \frac{1}{N} \log_2\left(p\left(r_j\right)\right). \tag{3.37}$$

Here, $p(r_j)$ is the local estimate of the probability density function at spike time r_j. Essentially, this has reduced the complexity of the problem by allowing one to estimate the probability density only at the observed spike times, as opposed to at all possible times.

A local estimate of the probability density function is arrived at by considering the distance of this spike time from its nearest neighbor, the next closest sampled spike time. Qualitatively, if the observed spike time r_j is very close to other observed spike times, then $p(r_j)$ should be high. However, if the observed spike r_j is an outlier, far away from other observations, $p(r_j)$ should be low. Formally, this relationship can be shown to be

$$-\log_2\left(p\left(r_j\right)\right) \approx \lambda_j + \log_2[2(N-1)] + \frac{\gamma}{\ln(2)}, \tag{3.38}$$

where λ_j is the time between spike r_j and its nearest neighbor and γ is the Euler-Mascheroni constant ($\gamma \approx 0.5772$).

Computing MI as a difference of entropies, H(R)-H(R|S) (Eqn. 3.7), the binless equation for MI_{timing} can finally be shown to be

$$MI_{binless} = \frac{1}{N}\sum_{j=1}^{N}\log_2\left(\frac{\lambda_j}{\lambda_j^*}\right) - \sum_{s\in\{S\}}\frac{N_{reps,1}(s)}{N}\log_2\left(\frac{N_{reps,1}(s)-1}{N-1}\right), \qquad (3.39)$$

where $N_{reps,1}(s)$ represents the number of times that stimulus *s* results in at least one spike (calculated as $p_{spike}(s)\ N_{reps}$), and λ_j^* is the time between spike r_j and the nearest spike elicited by the same stimulus. Thus, the timing information in the binless method breaks down to a comparison of the ratio between nearest neighbor spike distances without stimulus classification and the nearest neighbor spike distances with stimulus classification.

The binless method gives surprisingly accurate results even with limited amounts of data. Figures 3.14A, C, and E show how binless estimates of timing information converge as a function of sample size for the same simulations used to assess the accuracy of the KDE information measures. The KDE estimates are shown for reference. The bias in the binless method is lower than in the KDE method, although the variance is higher. Comparing the estimation errors of the two measurements (Figs. 3.14B, D, and F) shows that the binless estimates tend to converge faster than the KDE measures, although they are often less accurate for low numbers of stimulus repetitions.

The binless method suffers from one drawback for a study based on two stimulus variables. Because the probability density functions are approximated as a sum of delta functions, the marginal probability density functions, p(x,r) and p(y,r), are not directly derived from the joint probability distribution p(s,r). (Since s is a subset of x, the number of delta functions in p(x,r) is more than in p(s,r).) As will be shown in section 3.8, if

estimates of the marginal distributions are not precisely related to estimates of the joint distribution (as expressed in Eqn. 3.44) then the cue-interaction equation (Eqn. 3.13) will not hold. Thus, while the binless method can provide more accurate measures of overall first spike latency information than the KDE method, it cannot be used to estimate the relative information. If one wishes to know the percentage of the full information that is devoted to coding individual cues, one must use another technique.

3.8 Derivation of the cue-interaction equation, Eqn. 3.13

In this section, a derivation of Eqn. 3.13 expressing MI_{FULL} in terms of MI_X and MI_Y is given. Using the Chain Rule for information (Cover and Thomas, 1991), MI_{FULL} can be rewritten as

$$MI_{FULL} = MI(X,Y;R) = MI(X;R) + MI(Y;R \mid X), \tag{3.40}$$

where the conditional information is defined in terms of entropies as

$$MI(Y;R|X) = H(Y|X) - H(Y|R,X). \tag{3.41}$$

Using Eqn. 3.3 to substitute for H(Y|X) in the above equation yields

$$\begin{aligned} MI(Y;R|X) &= H(Y) - H(Y|R,X) - MI(X;Y), \\ &= H(Y) - H(Y|R) + H(Y|R) - H(Y|R,X) - MI(X;Y), \\ &= MI(Y;R) + MI(X;Y|R) - MI(X;Y). \end{aligned} \tag{3.42}$$

The stimulus sets used in this study are designed such that MI(X;Y)=0, since every combination of localization cues is presented an equal number of times. Ignoring this term, and substituting Eqn. 3.40 back into Eqn. 3.42 yields

$$MI(X,Y;R) = MI(X;R) + MI(Y;R) + MI(X;Y \mid R), \tag{3.43}$$

which is exactly the cue-interaction equation (Eqn. 3.13).

One subtlety that may not be immediately obvious is the following. In order to perform the entropy substitution step of Eqn. (3.42), the following equation must hold:

$$p(x,r) = \sum_{y \in \{Y\}} p(x,y,r). \tag{3.44}$$

In words, the marginal distribution must be the sum of the joint distribution over the other variable. While this may seem like a trivial requirement, in real life one never has a access to the real probability distributions and estimates of p(x,y) and p(x) are made, instead. Care must be taken to ensure that the estimates of p(x) and p(x,y) obey Eqn. 3.44 exactly.

3.9 Derivation of the first spike latency information equation, Eqn. 3.31

To understand the breakdown of the first spike latency information, it is helpful to realize that the response variable, **FSL**, is actually a vector made up of two components, **FSL**=(C,T), where C is a random variable representing whether or not a spike was observed and T represents the arrival time of the observed spike. Thus, the information between the stimuli and the first spike latency, MI(S;**FSL**), is actually MI(S;C,T). Using the Chain Rule for information (Cover and Thomas, 1991), one can show

$$MI(S;C,T) = MI(S;C) + MI(S;T|C). \tag{3.45}$$

Since the count can take on only one of two values, the above equation can be expanded to

$$MI(S;C,T) = MI(S;C) + p(C=0)MI(S;T|C=0) + p(C=1)MI(S;T|C=1). \quad (3.46)$$

Of course, there can be no information in spike timing when a spike is not observed. Hence, MI(S;T|C=0)=0, and the above equation collapses to

$$MI(S;C,T) = MI(S;C) + p(C=1)MI(S;T|C=1), \quad (3.47)$$

which is exactly the first spike latency information equation (Eqn. 3.31).

4. Sound localization information encoded in spike rate

Perhaps the most obvious neural code for stimulus parameters is spike rate, and most of the previous studies on sound localization coding discussed in the introduction were based on the assumption of a rate code. This chapter will focus on the rate representation of sound localization cues in IC neurons, with particular attention paid to the differences in representation across the physiological classes of neurons defined by Ramachandran and colleagues (1999). It is often thought that parallel pathways of the brainstem auditory system have arisen to process different aspects of acoustic information, as in the segregation of ITD and ILD related processing in the MSO and LSO, respectively. One question that this chapter will address is the extent to which different types of localization information remain segregated among the classes of IC neurons.

This chapter begins with an overview of the data in section 4.1, which discusses the numbers of each type of neuron encountered. The coding of the SN/ILD and ITD/ILD stimulus sets is described in section 4.2, with the coding of the ABI/ILD stimulus set occurring in section 4.3. In section 4.4, the major localization-coding results of the chapter are discussed. Finally, section 4.5 discusses the issue of non-stationarity and

introduces a new renormalization technique that allows one to correct for slow drifts in the average firing rate of the neuron. The majority of the results presented in this chapter have been published (Chase and Young, 2005).

4.1 Overview of the data

In order to make accurate estimates of MI, only neurons from which at least 10 repetitions of one or more stimulus sets were collected are considered. This corresponds to 149 neurons from 37 cats. A histogram showing the number of neurons that were held for various numbers of stimulus presentations is shown in Fig. 4.1.

Classification was done according to the Ramachandran *et al.* (1999) scheme, as described in section 2.3 of the Methods chapter. In sum, 119 of the 149 neurons were successfully classified, including 40 type-O cells, 50 type-I cells, and 29 type-V cells. The remaining neurons represent 10 onset neurons and 20 neurons with unclassifiable response maps. The complete breakdown of the neurons in terms of the numbers of each cell type recorded with each stimulus set is shown in Table 4.1.

For the results discussed in this chapter, all of the debiasing was done with the bootstrap procedure described in section 3.5.2 of the Mutual Information and Bias chapter. This procedure was chosen because (1) it shows low error values for the numbers of stimulus repetitions we typically achieved (Fig. 3.4) and (2) the residual bias converges to the true value in a monotonic fashion. This latter property helps ensure the stability of the relative information calculations.

4.2 Rate responses to the SN/ILD and ITD/ILD stimulus sets

4.2.1 The representation of single cues

Figure 4.2 presents two examples of the relationship between spatial cue information and spike count. The top row shows data for a 6 kHz type-I neuron investigated with the SN/ILD stimulus set. The spike rate is plotted as a function of the two stimulus parameters in Fig 4.2A. For this example, MI_{FULL} is 1.75 bits. The contours on the back walls show the mean and standard deviation of the spike rate when considering only the response to the individual cues, achieved by collapsing the surface along one axis (i.e., averaging across the other cue). Figures 4.2B and 4.2C replot these spike rate contours as a function of ILD and SN, respectively. This neuron is much more sensitive to changes in ILD than in SN, as MI_{ILD} is 1.29 bits, whereas MI_{SN} is only 0.21 bits. As indicated by Eqn. 3.13, MI_{FULL} does not equal the sum of the MIs carried about the individual localization cues. The bottom row presents SN/ILD data collected from another type-I neuron with a similar BF (5.1 kHz). This unit is more sensitive to changes in SN than to changes in ILD.

Figure 4.2 shows two type-I neurons of similar BF but with differing degrees of SN sensitivity. What determines this sensitivity? One difference between the two neurons is the sharpness of tuning. This sharpness is quantified with the Q_{40}, which is the neuron's BF divided by the range of frequencies around BF leading to excitation, as measured with contralateral tones presented at 40 dB above threshold. The neuron with low SN sensitivity (Fig. 4.2A) has a Q_{40} of 1.2, whereas the neuron with high SN sensitivity (Fig.

4.2D) is sharper, with a Q_{40} of 3.2. It should be emphasized that the changes in SN frequency are very subtle compared to the excitatory bandwidths of both of these neurons. As shown in Fig. 4.3, the SN center frequencies do not tend to stray from the excitatory portions of the response maps.

Figure 4.4 shows another example, this time of two type-O neurons in response to the SN/ILD stimulus set. The first neuron (Fig. 4.4A-C), with a BF of 10 kHz, shows very little rate sensitivity to SN cues, with an MI_{SN} of only 0.2 bits. The second neuron (Fig. 4.4D-F) has twice the SN sensitivity of the first, with an MI_{SN} of 0.4 bits. It is impossible to compute the Q_{40} for type-O neurons, as by definition their response to contralateral tones turns inhibitory at high levels. In fact, their excitatory response to broadband noise cannot be explained by their responses to pure tones, demonstrating the non-linear properties of this class. The differential sensitivity to SN cues for type-O neurons may be a reflection of the properties of their inputs (primarily from the spectral notch sensitive neurons of the DCN; Davis 2002).

To show that there are consistent differences among the responses dependent on cell type, examples of the normalized discharge rate profiles to the localization cues are plotted in Fig. 4.5. For every stimulus set, each neuron produced five rate profiles in response to each parameter. Because these five rate profiles often changed as a function of the second parameter, all five are included in this analysis. Rate profiles are normalized to their maximum discharge rate for comparison across neurons. Because in many cases there are too many data to see details in the plots, only the 40 rate profiles with the largest single-cue MIs are shown. Profiles corresponding to $MI_X < 0.1$ bits are not included to eliminate insensitive neurons.

The first column shows the rate profiles of the three cell types in response to ILD. Type-I units (middle) have the most consistent responses to the ILD stimuli, with almost all of the rate profiles showing a peak when the contralateral stimulus is louder, corresponding to sound sources in the contralateral hemi-field. Type-V units (top) usually show a peak at midline, where the overall binaural level is highest for these stimulus sets. Type-O units (bottom) show the most diversity in their responses, with both contralateral and ipsilateral selectivity. These results are in agreement with those described by Davis *et al.* (1999).

The second column shows the rate profiles of the three cell types in response to SN manipulations, plotted as a function of the center frequency of the SN in octaves re BF. Type-I neurons show a trough-shaped response centered on BF, where the notch is centered directly over the neuron's excitatory area. This profile is consistent with the type-I response map (Fig. 1.3B) in that the response is weakest when the SN is centered directly over the excitatory portion of the response map, and strengthens as the SN moves away from BF.

The type-V rate profiles show little modulation by SN. This is probably because of their wide excitatory bandwidths, which are not obviously truncated by inhibition; this point will be further discussed later in this section. The type-O neurons again show the most diversity in their rate profiles. In particular, there are a few peak-like responses to SN, some monotonically increasing responses, and a number of trough-like responses.

The third column shows the rate profiles in response to ITD manipulations. The ITD responses are quite variable across all cell types, though this diversity is most marked in

the type-O population. Indeed, of all the cell types, the type-O neurons seem to exhibit the most rate modulation in response to ITD cues.

The three cell types differ slightly in terms of MI_{FULL}, the information carried about the whole stimulus ensemble. Figure 4.6 shows MI_{FULL} plotted as a function of BF for all neurons studied with the SN/ILD and ITD/ILD stimulus sets. There is considerable overlap in MI_{FULL} across the cell types, especially at low BFs; however, for BFs above 4 kHz the type-I neurons carry slightly more information about these virtual space stimuli than the type-O neurons ($p<.01$, rank sum).

Figures 4.7A, 4.7B, and 4.7C show how each of the cell types codes for individual localization cues. In these cases, MI_X is plotted as a function of the BF of the neuron for X = ITD, ILD, and SN.

There is a strong BF dependence in the coding of the ITD cue (Fig. 4.7A), with low BF neurons having higher MI values than high BF neurons. Within similar BF ranges, however, it appears that type-I and type-O neurons code ITD as well as type-V neurons. Ramachandran and May (2002) have reported sensitivity to binaural tone beats in 10/10 type-V units, 15/22 type-I units, and 6/14 type-O units in the low BF (< 3 kHz) population. The present results show a larger scatter (especially among type-V neurons), but agree qualitatively with their results.

In the ILD data (Fig. 4.7B), there is complete overlap in the coding of MI_{ILD} across the cell types. Although type-I neurons encode ILD in a straightforward fashion (Fig. 4.5), they do not convey more information about ILD than do types V and O ($p_{I/O}=.26$, $p_{I/V}=.10$, rank sum). However, at this stage ILD sensitivity cannot be dissociated from an overall level sensitivity. This point will be addressed in section 4.3, where the sensitivity

of type-V neurons to ILD is shown to be smaller when ABI is controlled. The largest difference among the populations is that there is a significant negative correlation between MI_{ILD} and BF for the type-V population ($r=-0.41$, $df=25$, $p<.05$), while there is no significant correlation with BF for the type-I or type-O populations ($p>.1$ for both).

The coding of SN (Fig. 4.7C) exhibits some BF dependence, with higher BF neurons showing more sensitivity than lower BF neurons. This corresponds to the fact that cat HRTF spectral notches are restricted to the range 6-20 kHz (Musicant *et al.*, 1990; Rice *et al.*, 1992) and suggests that neurons in the IC may be tuned to extract spectral notches where they are most likely to occur, physiologically. In fact, information about SN is found mainly in a subset of the type-O and type-I populations, in that only 4/13 type-V neurons have MI_{SN} greater than 0.2 bits, as opposed to 22/40 type-I, and 15/27 type-O neurons. No type-V neurons have MI_{SN} greater than 0.5 bits. By comparison, less than 20% of neurons from the SN/ILD set (3/13 type V, 3/40 type I, and 8/27 type O) have MI_{ILD} less than 0.2 bits. The highest values of MI_{SN} achieved in these experiments were in type-I neurons. The difference between the type-I and type-O populations with BFs above 4 kHz is small, not significant at the .05 level (mean, type-I: 0.5 bits; mean, type-O: 0.3 bits; $p=.15$, rank sum).

Receptive field width can be a good predictor of SN sensitivity in some neurons. Figure 4.8 shows a plot of MI_{SN} as a function of Q_{40} for those neurons with a well-defined excitatory area to tones 40 dB above threshold. The SN information is positively correlated with Q_{40} ($r=.66$, $df=39$, $p<1e-6$), explaining why type-V neurons do not carry much SN information in their firing rates.

4.2.2 The representation of multiple cues

By presenting stimuli that vary in two stimulus features simultaneously, the relative sensitivity of the neuron to those features can be evaluated. Figure 4.9 shows plots of the fraction of MI_{FULL} devoted to SN versus ILD (Fig. 4.9A) or to ITD versus ILD (Fig. 4.9B); the axes plot the relative MI, MI_X/MI_{FULL}, as a percentage. This way of plotting the data also allows the effects of differences in the total information encoded by a neuron, due to experimental effects like the sound level of the stimuli relative to the neurons' dynamic ranges, to be normalized.

The diagonal black lines in Fig. 4.9 show the expected behavior of the data if MI_{FULL} were the sum of the MIs to the two stimuli separately. Deviations from these lines are the confounded information derived in Eqn. 3.43. There is confounded information because the response variable, rate, is unidimensional and is shared by the responses to two cues. Confounded information is maximal for neurons that encode the two stimuli approximately equally. It is minimal when one stimulus dominates the neuron's responses, at the upper left or lower right corners of the plots.

For SN versus ILD information (Fig. 4.9A), there is a range of behaviors, with many neurons dominated by ILD (lower right corner). There are no signs of coding specialization within the type-O and type-I populations, in that neurons of both types distribute across the whole range of relative sensitivity. Type-V neurons, however, never have more than 50% of MI_{FULL} accounted for by MI_{SN}.

The tradeoff between ITD and ILD coding is shown in Fig. 4.9B. Neurons with BFs greater than 4 kHz are not shown, as ITD sensitivity is weak for high frequency neurons and these points all cluster in the ILD corner of the plot. Among low frequency neurons, there is a range of ITD coding sensitivity for all cell types. Type-V neurons show significant ITD coding, but the other cell types are represented there as well. Note that some of these low-BF neurons are sensitive to ILD cues in the absence of ITD sensitivity (lower right corner of the plot), despite the small size of physical ILD cues at low frequencies (Irvine, 1987).

4.3 Responses to the ABI/ILD stimulus set

4.3.1 Information about overall level vs. level differences

The ABI/ILD stimulus set was designed to help disambiguate true ILD coding from monaural or binaural sound-level coding. In this set, ILD varies as ABI is held constant, and vice-versa. The normalized rate profiles are plotted in Fig. 4.10, to show the differences in the form of responses to these cues across cell types.

The type-V neurons (top row) display consistent increases in rate with increasing ABI and a diverse range of response to ILD. This result implies that the information provided about ILD by type-V neurons in their responses to the SN/ILD and ITD/ILD stimulus sets was mainly a response to overall stimulus level, as opposed to a tuned response to midline azimuths as implied by Fig. 4.5. For both stimulus sets shown in Fig. 4.5, ABI

peaks at the midline. The type-V rate-profile results agree with the hypotheses of Ramachandran *et al*. (1999) and are consistent with the EE nature of type-V neurons.

The type-I rate profiles (second row) show a clear contralateral hemi-field preference when using the ABI constant method, though in a few instances there are mildly tuned responses to midline azimuths. This result is generally consistent with the data in Fig. 4.5. Most of the cases studied here also show an increase in response with increasing ABI.

The type-O population again shows the most diverse rate profiles. While the majority of neurons show a contralateral hemi-field preference to ILDs, as in Fig. 4.5, some neurons show a tuned ILD response, and in some cases there are ipsilateral hemi-field preferences. The response to increasing ABI shows no consistency in this population.

The representation of ABI-constant ILD cues is shown in Fig. 4.11A. When ABI is held constant, type-I neurons carry significantly more ILD information than type-V neurons ($p<.01$, rank sum). Conversely, MI_{ABI} (Fig. 4.11B) is higher in type-V neurons than in type-I neurons ($p<.05$, rank sum), and type-V neurons have a larger ABI% (Fig. 4.11C) than type-I neurons ($p<0.01$, rank sum). The difference between type-V and type-O neurons is not significant in any of the three figures, though in each case the type-O neurons distribute more similarly to the type-I neurons than to the type-V neurons.

The essentially binaural nature of the responses is demonstrated by the fact that MI_{ABI} is generally small (mean 0.37 bits), most often less than 25% of MI_{FULL} (Fig. 4.11C; most of the exceptions are type-V neurons). By contrast, most MI_{ILD} measurements are larger (mean, 0.79 bits, Fig. 4.11A), with values comparable to those given in Fig. 4.7B. If neurons were truly monaural, MI_{ABI} and MI_{ILD} would be equal.

4.3.2 Assessing monaural vs. binaural interactions: the Monaural Index

When ILD is manipulated with either ITD or SN, it is impossible to disambiguate whether the neuron is responding to true level differences, or whether the response is mainly determined by the monaural level manipulations in one of the ears. When ILD is manipulated along with ABI, however, these effects can be distinguished. For this purpose, we developed an index that quantifies the extent to which a neuron's response is dominated by level manipulations in one ear verses both ears. This index is called the Monaural Index.

The Monaural Index compares the fraction of variance in the rate response accounted for by the contralateral sound level (fv_c) to the fraction of variance accounted for by the ipsilateral sound level (fv_i). If $r_{i,(j,k)}$ is the i^{th} response to the stimulus with contralateral level = j and ipsilateral level = k, then fv_c is calculated as

$$fv_c = 1 - \frac{\sum_j \sum_i \left(r_{i(j,0)} - \mu_j\right)^2}{\sum_j \sum_i \left(r_{i(j,0)} - \mu\right)^2}. \qquad (4.1)$$

Here, μ represents the neuron's mean rate response to all stimuli with ipsilateral level = 0 regardless of contralateral level, and μ_j represents the mean to all stimuli with contralateral level = j and ipsilateral level = 0. There are five such stimuli in the ABI/ILD set, with levels (16,0), (8,0), (0,0), (-8,0), and (-16,0), where the values in parentheses are the contralateral and ipsilateral levels, respectively. fv_i is calculated in a similar manner, but from a different set of stimuli with the contralateral level held

constant at 0 and the ipsilateral level allowed to vary. With these values, the Monaural Index is defined as

$$\text{Mon Ind} = \frac{\text{fv}_c - \text{fv}_i}{\text{fv}_c + \text{fv}_i}. \tag{4.2}$$

When the Monaural Index is equal to one, none of the variance in the rate is accounted for by ipsilateral level; the neuron is monaural to the contralateral ear. Similarly, when the Monaural Index is equal to negative one, the neuron is monaural to the ipsilateral ear. When the Monaural Index is zero, the rate variance accounted for by contralateral level exactly matches that accounted for by ipsilateral level, i.e., the neuron is fully binaural. Note that a binaural unit can be sensitive to ABI or to ILD, and in either case the Monaural Index can go to zero. In this case, ABI verses ILD sensitivity can be distinguished using the mutual information.

To explore how monaural or binaural sensitivity might contribute to the segregation of ABI and ILD coding, a histogram of the Monaural Indices computed for all the neurons measured with ABI/ILD stimuli is plotted in Fig. 4.12A. As expected, this distribution is skewed towards positive values (contralateral sensitivity). Monaural Index values of less than –0.2 occur only for neurons with BFs ≤ 3kHz.

Surprisingly, the type-O population turns out to show the most binaural sensitivity, with a tight cluster of responses around zero. The type-V population, on the other hand, shows the least binaural sensitivity, having the most Mon-Inds that exceed 0.5 in absolute value. The absolute value of the Monaural Index can be used as a measure of the degree of binaural level interaction (with 0 being the most and 1 being the least). With this measure, the type-V population has significantly less binaural interaction than either the type-I or type-O populations ($p<.05$, rank sum).

As expected, MI_{ILD} is largest in neurons with values of Monaural Index near zero (Fig 4.12B). While MI_{ABI} (Fig. 4.12C) does not show much correlation with Monaural Index, MI_{ILD} decreases almost linearly with the log of Monaural Index. With one exception in the type-V category, binaural neurons code ILD information much better than ABI information.

4.4 Discussion of rate coding results

4.4.1 Segregation of localization cue information among the IC cell types

As per the discussion in section 1.2.4 in the Introduction, sound localization cues appear to be largely segregated in the auditory brainstem. Although an information analysis of brainstem neurons has not been done, present results suggest that localization cues are less segregated in the IC than in the brainstem nuclei. Given the hypothesis that IC response types are dominated by input from distinct brainstem nuclei, one expects to see similarities between LSO and type-I units, DCN and type-O units, and MSO and type-V units. The limited segregation of localization cues expected on the basis of this hypothesis is not observed.

There is complete overlap in the type-O and type-I populations in the coding of ILD cues, despite weak ILD coding in the DCN and strong ILD coding in the LSO. SN cues are represented similarly in the type-I and type-O populations, and in fact there is slightly more information about SN in the type-I population, especially at BFs appropriate to notch sensitivity. This result is contrary to the suggestion that the DCN and type-O

neurons form a specialized channel for SN (Young and Davis, 2002; Davis et al. 2003). Both type-I and type-O units encode ITD cues, which would not be expected from the characteristics of DCN or LSO neurons. (Note: there is an important caveat to this interpretation, and that is that MI measures neural sensitivity, not selectivity. This point is discussed in detail in section 7.3.)

Unlike type-I and type-O units, type-V units show properties similar to their hypothesized input from MSO neurons. They are sensitive to ITD and ABI (Goldberg and Brown, 1969) and relatively less sensitive to ILD. They are not sensitive to SN, but low BF neurons in general do not provide much rate information about SN. One difference from the hypothesized MSO input is that of all cell types, type-V neurons are most often dominated by monaural input, as indicated by the histogram of Monaural Indices (Fig. 4.12A).

In sum, these results suggest substantial convergence of information across ascending pathways within the IC, especially for the type-I and type-O classes. This is consistent with the finding that type-O units require a convergence of inputs with diverse properties to explain their responses to noise notches (Davis *et al*. 2003), even though they have been shown to depend on inputs from the DCN for their primary properties (Davis 2002).

It is possible that other classification schemes might result in a better segregation of localization cue information. However, the majority of units in this study respond to multiple cues, as shown by the percentage of units in Fig. 4.9 that cluster towards the center of the plots. It is difficult to envision a classification scheme that would segregate these responses.

These results also suggest that studies aimed at the integration of localization cues in the IC will be profitable. The process of constructing the spatial location of sounds relies on varying degrees of cue integration, depending on conditions (e.g. Wightman and Kistler, 1992). In addition, particular cues may be useful in solving various discrimination problems: ITD cues are predominant in binaural unmasking and SN cues are essential for elevation and front/back disambiguation. The relative reliability of ILD versus ITD cues depends on the frequency spectrum of the stimulus and the degree of reverberation. It is likely that the auditory system requires different degrees of convergence of localization cues in different conditions; the existence of neurons with widely varying degrees of cue integration, suggested by Figs. 4.9 and 4.11, would allow adaptation to different listening conditions and tasks.

4.4.2 Matching stimulus cues to neurons

Accurate estimation of MI requires a large amount of data, which restricts the number of possible stimulus parameters that can be tested. Because of the sparse sampling of the stimulus space, it is possible that the sensitivity of narrowly tuned neurons can be underestimated.

The localization cues tested in this study match the average receptive field widths of IC neurons. In general, neurons in the IC have fairly broad spatial receptive fields, with azimuth half-widths for tuned units averaging about 60° (Aitkin and Martin, 1987; Delgutte *et al.*, 1999; Sterbing *et al.*, 2003). Furthermore, ITD discrimination thresholds for single neurons in the IC of guinea pigs have been estimated to range from tens to

hundreds of microseconds (Shackleton *et al.*, 2003), the range tested in this study. While elevation *per se* was not tested, the SNs used represent elevations ranging in 7.5° steps from 0° to 30°. Units in the IC that show elevation tuning tend to have half-widths greater than 20° (Aitkin and Martin, 1990; Sterbing *et al.*, 2003). The range of parameters used here spans these receptive fields.

4.4.3 Comparison with other studies

Among studies that classified IC units in the same way as this study, Davis *et al.* (2003) looked at the processing of SNs in the type-O pathway. They found that the majority of type-O units showed an excitatory peak when a SN was slightly below BF. In contrast, only some of the type-O units in this study show an excitatory peak to the SN cue, with about half exhibiting a trough-like response. One possible explanation for this discrepancy is the limited sampling of SN center frequencies, as the Davis study sampled notches in a 2-octave range around BF, compared to 0.2 octaves for this study. It is unlikely that this fully explains the differences, however, as the stimuli used here differed enough to produce significant rate modulations. Another possible explanation is that their study used straight-walled, full-depth notches embedded in otherwise white noise, whereas this study employed more natural HRTF-filtered white noise.

Tollin and Yin (2002) showed that LSO azimuth tuning is primarily determined by the ILD in a narrow frequency band near BF. In contrast to the results presented here for type-I neurons, spectral shape did not determine a significant component of the responses in their study. This apparent difference probably derives from differences in the way SN

stimuli were presented. Here, the spectra were shifted to center the SN frequency range on BF, whereas Tollin and Yin used natural SN frequencies, which only occasionally aligned with BF.

4.5 Renormalization: correcting for slow drifts in neural excitability

In building the statistics necessary to make estimates of MI, the stimulus set must be presented multiple times; typically, recording sessions last on the order of an hour. Central to these estimates is an implicit assumption that the behavior of neurons remains stationary on this time scale. Studies of data from the auditory nerve (Teich, 1989) and lateral superior olive (Teich *et al.*, 1990), however, suggest that neurons in the auditory system may exhibit long-term correlations in their firing rates. This section describes one method of correcting MI estimates for slow drifts in the firing rates of single neurons. This correction is found to lead to a slight, but statistically significant, increase in the MI carried by spike rate about localization cues.

Figure 4.13 shows two examples of neurons exhibiting slow drifts in firing rate across multiple repetitions of a stimulus set. The blue circles denote the average firing rate to the entire block of 25 stimuli in the set as a function of the number of stimulus sets presented. In both examples, slow drifts in the average firing rate over time are evident. It should be noted that in each example, the isolation and triggering of the neuron was excellent.

To correct for non-stationarities in the response, the firing rates for each of the 25 stimuli within a block were normalized by the average firing rate to that block. This

forces the average firing rate across blocks to be the same, while still allowing variation in the relative rate to each stimulus.

Figure 4.14A shows the average firing rate (± 1 SD) as a function of stimulus for the example neuron of Fig. 4.13A. The firing rate is well modulated by the localization cues, leading to a rate MI of 1.62 bits. Figure 4.14C shows the renormalized firing rate as a function of stimulus for the same neuron. The y-axis limits were chosen to allow comparison of the size of the error bars between the Fig. 4.14A and 4.14C. A careful look at the two plots reveals that the standard deviation of the renormalized firing rate to each stimulus is smaller than in Fig. 4.14A.

If the firing rate process were stationary, normalizing the spike rate by the average firing rate within each block of stimuli would be equivalent to normalizing the spike rate to the average firing rate across all of the blocks of stimuli. Figure 4.14E shows the expected standard deviation of the firing rate to each stimulus under the stationary assumption, derived by dividing the standard deviations in 4.14A by the average firing rate across all repetitions of all stimuli (blue circles). For comparison, the actual renormalized standard deviations are also shown (red asterisks). Clearly, the standard deviations of the renormalized rates are smaller than the expected standard deviations given a stationary process. Renormalizing the firing rates on a block-by-block basis leads to a tighter distribution of firing rates to each stimulus.

The MI between the full set of localization cues and the renormalized firing rates can be computed in the same fashion as the MI between spike count and the localization cues, with one exception. When the firing rates are normalized by the average firing rate to each block, the discrete distribution of spike counts becomes a continuous distribution of

rates. To compute the MI in this case, the renormalized rate variable was divided into a number of equally spaced bins, with the probability of each bin determined by the number of samples that fell into it. The number of bins was set equal to the number of different spike counts observed, such that the total number of response variables was the same between the renormalized rates and the spike counts. Since bias is primarily a function of the number of stimulus and response variables (see section 3.5.2), this choice in the number of bins helps ensure that the biases for the spike count MI and the renormalized rate MI are consistent.

The tighter distribution of spike rates observed in Fig. 4.14C leads to a larger MI_{FULL} for the renormalized rates than for the spike counts alone (1.91 bits as opposed to 1.62). Figures 4.14B, D, and F show similar results for the example neuron of Fig. 4.13B.

In general, renormalizing the firing rates leads to increases in information. Figure 4.15 shows the renormalized MI plotted as a function of the spike count MI for all of the neurons in the population. The two example neurons of Figs. 4.13 and 4.14 are shown outlined in red boxes. Most of the neurons exhibit a rate increase when the spike counts are renormalized, as shown by the proportion of neurons above the diagonal. The inset histogram shows the number of neurons that exhibit various renormalization gains; defined as the difference between the renormalized MI and the spike count MI. The histogram is significantly skewed toward positive values ($p<1e-20$, t-test). However, the mean value of the information increase is modest, at 0.1 bits. Thus, the assumption of stationarity does not lead to hugely incorrect estimates of MI.

5. Sound localization information recovered by a decoder

Spike count represents a very coarse measure of neural activity. In the auditory brainstem, neurons tend to fire action potentials in response to naturalistic stimuli at well-defined times relative to the stimulus. Multiple repetitions of these stimuli can lead to very precise bursts in the PSTH, the locations of which often change with manipulation of stimulus parameters. Thus, it is possible that not only the numbers of spikes during a stimulus carry information, but also the locations of those spikes within the stimulus presentation period.

This chapter investigates the coding of multiple localization cues in the spike trains of single IC neurons using a stimulus decoder. The decoder, which is based on a spike distance metric (SDM) developed by Victor and Purpura (1997), allows the estimation of the MI between localization cues and the spike trains at several different temporal resolutions. This approach allows one to separate out the contributions to information arising from spike timing and spike rate.

Section 5.1 describes the spike distance metric and how it can be used to make estimates of MI. In section 5.2, the temporal coding of individual localization cues is

considered, while in section 5.3, the coding of multiple cues is investigated. Section 5.4 concludes with a discussion of the major results of this chapter. The majority of the results presented in this chapter have been published (Chase and Young, 2006).

5.1 The spike distance metric decoder

The SDM decoder (Victor and Purpura, 1997) works by developing a notion of similarity between spike trains. Intuitively, if two spike trains are similar, it is more likely that they represent the same stimulus than two spike trains that are different. The SDM formalizes the concept of similarity between spike trains by defining a distance between spike trains. Essentially, the distance between two spike trains is defined as the sum of the costs of the elementary steps it takes to transform one spike train into the other. The allowed steps are spike deletion (cost of 1), spike addition (cost of 1), and spike shift (cost of $q|\Delta t|$), where Δt is the time difference between a spike in one train and the nearest spike in the other train and q is a variable cost parameter (units of s^{-1}). For a given q, there exists a minimum cost solution for the distance between any two spike trains. A schematic of the distance calculation is presented in Fig. 5.1A.

The cost parameter q represents the precision with which spikes are timed. If $q|\Delta t|>2$, it is cheaper to add and delete a spike than it is to shift it. Thus, $1/q$ is proportional to the time interval between spikes at which they are considered to be different, which can be interpreted as the integration time of a neuron reading the spike train. If $q=0$, the only distance assigned between spike trains is the difference in their absolute number of spikes, which represents a rate code. At the other extreme, as q approaches infinity, the

reading neuron performs a coincidence detector function where spikes are not considered associated unless they occur at exactly the same time.

With the notion of spike train distance defined, for any given spike train in response to stimulus *i*, it is possible to calculate the average distance to every other spike train measured from that neuron elicited by stimulus *i*, $\langle d(i,i) \rangle$. This average distance can then be compared to the average distance between the spike train and every spike train elicited by stimulus *j*, $\langle d(i,j) \rangle$. Figure 5.1B illustrates this idea, where the dots represent spike trains and colors represent the stimuli being presented when the spike trains were measured. To a reading neuron, dots that cluster closest together should produce the most similar responses. To compute the information between stimuli and spike train distances, the spike trains are assigned to the groups to which they are closest, regardless of the actual stimulus. That is, the spike train *i* is estimated to have come from stimulus *j* when *j* satisfies:

$$\left\langle d(i,j) \right\rangle < \left\langle d(i,k) \right\rangle \quad \forall \ \mathrm{k} \neq \mathrm{j}. \tag{5.1}$$

After repeating this process for every spike train, a confusion matrix *N* is created where $N(i,j)$ represents the number of times a spike train from stimulus *i* is classified as being closest to spike trains from stimulus *j* (Fig. 5.1C). The confusion matrix, when normalized by the total number of stimulus presentations, defines the joint stimulus/response probability on which MI is calculated, as described in section 3.6 in the chapter on MI. As a stimulus estimation technique, the SDM allows the computation of a lower bound on the MI between stimuli and responses in a way analogous to other decoding techniques (Kjaer *et al.*, 1994; Rolls *et al.*, 1997; Furukawa and Middlebrooks, 2002).

Since the distance metric is a function of q, the MI calculated with this method is also a function of q. In addition to the q=0 case, q was set to range from 10 to 15850 s^{-1} (100 to 0.063 ms), sampled logarithmically at 5 costs per decade. These costs were found empirically to cover the relevant range of timing resolutions of the IC neurons studied.

Because of the added difficulty in making accurate confusion matrices with limited numbers of samples, only neurons with at least 20 repetitions of a stimulus set are considered in this chapter. The bootstrap technique, performed on the counts of the confusion matrix as described in section 3.6.2, was used to debias the MI estimates.

5.2 SDM analysis of the coding of individual localization cues

5.2.1 Single-cue information

Figure 5.2 shows an example of a neuron studied with the ITD/ILD stimulus set presented at a high sound level. As often happens at high levels, the rate response is saturated (Fig. 5.2A), so the rate information about the full stimulus set (rMI_{FULL}) is only 0.3 bits. (For the remainder of this chapter, MI computed directly from spike count will be referred to as rMI.) Though there is little consistent change in the spike count among stimuli, a close up view of the spike rasters (Fig. 5.2B) shows considerable variation in individual spike times with ILD. In particular, whereas the first burst is either on or off depending on the ILD, the second, third, and fourth bursts are progressively delayed with increases in ILD. Figure 5.2C shows the results of the SDM analysis on these spike trains. For a spike-shift cost of 1000 s^{-1}, 1.6 bits of information is recovered about the

stimulus identity. This maximum is called MI_{peak}, and the cost at which it occurs is called the peak cost. The cost=0 case, which represents a rate code, is called MI_0. Finally, the largest cost at which the MI decays to half of its peak value is known as the cutoff cost, which is ~4000 s^{-1} for the neuron in Fig. 5.2C. The MIs to the individual location cues are shown in Figs. 5.2D and 5.2E. As expected from the raster plot, most of the information in MI_{FULL} is about ILD.

Figure 5.3 shows another example; in this case there is little extra information available in spike timing that is not available in rate. The rate surface of Fig. 5.3A shows considerable variation in response to both stimulus parameters, and indeed rMI_{FULL} is quite high at 2.3 bits. The *MI*(*cost*) curve from the SDM analysis shows a nearly low pass behavior (Fig. 5.3C), with only a small peak that would indicate extra information available in spike timing. Note that this is not because this neuron does not exhibit stimulus locking, as shown in the raster plot (Fig. 5.3B). Rather, the variation in spike timing across stimuli is not significant compared to the rate differences.

The results in figures 5.2 and 5.3 exemplify the range of behavior shown by the population; typically, responses lie between these two extremes. The information carried in spike patterns about the full stimulus set, as assessed with the SDM method, is shown as a function of BF in Fig. 5.4A for all neurons in this study. To assess possible differences across groups, an ANOVA calculation was performed with a significance criterion of p=0.05, corrected for multiple comparisons. Frequency was divided into 3 equally populated groups (low, middle, and high) to assess the effects of BF as an independent variable. There are no differences in the MI_{peak} values across the three

stimulus sets. There are also no differences in MI_{peak} across the neuron types, nor differences in MI_{peak} values with BF.

To view the amount of information available in temporal spike patterns that is not available in spike rate, MI_{peak} from the SDM calculation is plotted as a function of the information calculated assuming a rate code in Fig. 5.4B. The vertical offset of the points from the diagonal represents the extra information available when spike timing is taken into account. Many neurons show a considerable information gain with the SDM method.

Recall that for the zero cost ($q=0$) analysis, no penalty is assigned to shifting spikes, only to adding or deleting them. This case should, then, correspond to a rate code, and the information calculated from the SDM method at zero cost should be the same as the information calculated under the assumption of a rate code, if no information is lost in the decoding step when the confusion matrix is generated. When MI_0 is plotted as a function of *rMI* for each neuron, there is very good agreement between the two measurements (r=0.99; Fig. 5.4C).

In Fig. 5.5, MI_{peak} values are compared to the corresponding *rMI* measures for each of the individual localization cues. For ITD information (Fig. 5.5A), the points all cluster along the diagonal, showing that very little extra information is available when considering spike timing. This strongly suggests that at the level of the IC, ITD information is carried in a rate code. The coding of ILD cues is shown in Fig. 5.5B, where ILD cues from all three stimulus sets have been lumped together, as the populations overlap. The SDM method recovers a mild amount of information about ILD cues over that available through a rate code. The same holds true for ABI cues (Fig.

5.5D), which show less information, on average, than the other cues. The largest effect of spike timing is on the coding of SN cues (Fig. 5.5C). In general, the MI in SDM responses to SN is much larger than the MI in rate responses to SN. For some neurons, as much as 2 bits of information is recovered by considering spike timing.

To investigate the nature of the SN timing information in more detail, the peak SN information computed with the SDM method is plotted as a function of BF in Fig. 5.6A. There does not seem to be much BF dependence to the MI_{SN} values, which is surprising given the BF dependence observed for rate coding of SN (Fig. 4.7C). To illustrate this point further, the spike timing gain for SN information is plotted as a function of BF in Fig. 5.6B. This gain is defined as the difference between MI_{peak} and MI_0, and represents information only available through spike timing. The spike timing gain is negatively correlated with BF ($r=-.46$, $df=78$, $p<0.0001$); it is mainly the low BF neurons that carry the extra information about SN in the timing of spikes, though there are some mid-frequency neurons with spike-timing gains of as much as 1 bit. The largest spike-timing gains are seen in type-V neurons, which are found only at low BFs (Ramachandran et al., 1999) and are the predominant low BF response type in our sample. This point is discussed in more detail in section 5.4.2.

5.2.2 Time scale of information

The cost at which MI is maximum is a measure of the temporal precision of the spike patterns that provide information captured by the SDM analysis. As discussed in Methods, $1/q$ is a measure of the effective integration time of a neuron reading the

temporal information in spike patterns, in the sense that $2/q$ is the maximum time delay between spikes in two trains at which they can still be shifted into alignment.

Figure 5.7 shows data on the costs at which the maximum MI is obtained with the SDM analysis, for each of the localization cues studied. A peak cost of zero was declared if the MI_{peak} value remained within 10% of MI_0 for that neuron (for example, the ILD curve in Fig. 5.3C), indicating that most of the information was carried by rate. Peak costs of zero occurred in 40% of the ITD cases (Fig. 5.7A) and 43% of the ILD cases (Fig. 5.7B). This is in comparison to only 9% of cases for the SN cue (Fig. 5.7C) and 13% of cases for the ABI cue (Fig. 5.7D). The median peak cost value obtained by pooling cost values from all localization cues and ignoring zero values is ~100 s^{-1}. Thus, when there is localization information in spike timing, it is integrated on a time scale of approximately 10 ms.

The major difference between the information time scales for different localization cues is that there is a significant correlation between BF and peak cost for the SN cue ($r=-0.43$, $df=66$, $p<0.001$, ignoring zero cost values) that is not seen with the other cues. SNs are represented at finer time scales in low-BF neurons than they are in high-BF neurons.

5.2.3 Frozen vs. random noise

In this section, we show that the information that is recovered by the SDM analysis is almost entirely derived from locking to temporal features of the stimulus. To demonstrate this point, responses to frozen noise, for which the temporal waveform is the same in all stimulus repetitions, were compared to responses to phase-randomized noise,

for which the temporal waveform differs in each repetition. Information that depends on the temporally-locked stimulus features will not be present with the random noise. All data presented to this point were obtained with frozen noise.

Figure 5.8 shows the responses of a 1.8 kHz type-O neuron in response to the random/frozen stimulus set described in section 2.4 in the Methods chapter. There is very little difference in the average rate responses, as shown in Figs. 5.8A and 5.8B, and the rMI_{FULL} of these stimulus sets are nearly identical at ~0.65 bits. The temporal responses to the two stimulus sets are completely different, however, as shown by the raster plots in Figs. 5.8C and 5.8D. From the raster to the frozen noise, it is clear that this neuron responded to specific temporal events in the stimuli, events that occur at fixed times in the frozen noise, but not in the random noise. As an example, there is a spike that occurs frequently at a latency of about 56 ms in the responses to two of the frozen SN stimuli (1.8 and 1.7 kHz), but not in the others. The only apparent temporal feature that remains in the random noise is the latency of the first burst of spikes, which changes systematically with ILD in both stimulus sets.

The SDM information carried by this neuron is shown in Fig. 5.9A-C for both stimulus sets. As is characteristic of low frequency neurons, there is a large peak in the *MI(cost)* curves for the frozen noise set, indicating the presence of a substantial amount of information in spike timing over that in rate. These peaks are missing from the random waveform responses. Thus, the extra information available in spike timing was due to the differences in the temporal waveforms within the frozen noise SN/ILD stimulus set, as opposed to an intrinsic variation in spike patterns stemming purely from spectral or level differences.

ILD-related latency differences were observed in both the responses to frozen and random waveforms (Figs. 5.8C and 5.8D). However, the SDM method reveals timing information in only the frozen waveform case. This indicates that the SDM method, as computed here, is relatively insensitive to first-spike latency variation. While differences in spike latency must increase the distance between spike trains, this distance is apparently overwhelmed by other noisy sources of spike train differences. When the SDM MI is computed using spike trains with all but the first spike removed (a first-spike latency code), 1.3 bits and 0.9 bits of information are recovered about the full stimulus set for the frozen and random waveform sets, respectively. Information in first spike latency will be discussed further in Chapt. 6.

The results of the example neuron in Figs. 5.8 and 5.9 are consistent across the population of neurons for which the random-waveform stimulus data were gathered. Figure 5.10 shows the spike-timing gain (MI_{peak}-MI_0) for the random noise stimulus set plotted as a function of the timing gain for the frozen stimulus set. The data cluster along the x-axis, indicating that no neurons exhibit much extra information in spike timing for the random noise stimuli. For the 14 neurons studied with the random noise stimuli, the mean and standard deviation of the spike timing gains for frozen noise were 0.51 ± 0.60 bits (range: 0 to 1.7 bits), while the corresponding spike timing gain values for random noise were significantly less at 0.09 ± 0.06 bits (range: 0 to 0.24 bits; different from frozen noise at $p<0.01$, signed rank test). ILD random noise spike timing gains were not significantly different from SN gains.

5.3 SDM analysis of the coding multiple localization cues

As seen in the last chapter, single neurons in the IC often respond to multiple stimulus features. Depending on how these features are coded in the neural response, the brain may or may not have access to each individual feature. This leads to a natural question: how can multiple stimulus features be multiplexed onto the responses of single neurons? As will be discussed below, the added dimensions of spike timing potentially allow for multiple-feature representation.

To examine the mixture of cues being coded at different time scales, the relative information can be computed as a function of the spike-shift cost, q. An example of this computation is given in Fig. 5.11 for a type-V neuron in response to the ITD/ILD stimulus set. It is clear from the rate surface shown in Fig. 5.11A that this neuron responds to both ITD and ILD manipulations, and the rMI is substantial at 1.2 bits. Fig. 5.11B reveals that there is a modest amount of extra information contained in spike timing about the full stimulus set. The MI(cost) curves for the individual localization cues, however, show that there is a distinct tradeoff between ITD and ILD information as a function of cost.

This information tradeoff is further emphasized in Fig. 5.11C, which plots the proportion of the full information being devoted to each cue as a function of cost. The relative MI_{ITD} shows a low-pass behavior, whereas the relative MI_{ILD} increases at higher temporal resolutions. The cost 'trajectories' are shown in Fig. 5.11D. Each dot plots the relative information for ITD versus ILD at a particular cost. The 0, 100, and 1000 Hz cost points have been labeled to indicate the direction of this cost trajectory. At low costs (representing a rate code), the neural response is dominated by the ITD information.

However, as the cost increases and the spike trains are analyzed at higher temporal resolutions, the ILD information dominates the response.

This neuron was held long enough to present the SN/ILD stimulus set as well; the responses to this set are shown in Fig. 5.12. The rate response surface (Fig. 5.12A) indicates that not only ILDs and ITDs drive the response of this neuron; SN changes also affect the firing rate. Figure 5.12B shows MI_{FULL}, MI_{ILD} and MI_{SN} as a function of cost. At ~630 s^{-1}, there is a prominent peak of 3.5 bits in the $MI_{FULL}(cost)$ curve (one of the most sensitive neurons in the population), a gain of about 2 bits over its zero-cost value. While it is clear from Fig. 5.12B that MI_{ILD} and MI_{SN} co-vary with MI_{FULL}, Fig. 5.12C shows that the fraction of MI_{FULL} devoted to ILD or SN coding is not constant with cost. Instead, there is a monotonic increase in the MI_{SN} percentage as a function of cost, while in this case MI_{ILD} shows a low-pass behavior. The three localization cues studied in this neuron show a clear segregation in their neural representation at different time scales. ITD cues are coded at the coarsest temporal resolutions (rate), ILDs are coded at intermediate temporal resolutions, and SNs are coded at the highest temporal resolutions.

The coding trajectory for this neuron, shown in Fig. 5.12D, reveals another property of spike-timing codes. Points lie below the diagonal when MI_{FULL} is not equal to the sum of the information in the individual cues, or when the confounded information of Eqn. 3.13 is non-zero. At a cost of 0 (rate case), the confounded information is large; MI_{SN} and MI_{ILD} are not independently represented in the neural response. As the cost parameter is increased, the confounded information decreases until it reaches zero (near 50 s^{-1}), signifying that SN and ILD are independently coded. Finally, at costs over 1000 s^{-1} the

information about both ILD and SN (and MI_{FULL}) decreases. The decrease is faster for ILD, so the points in Fig. 5.12D move toward the upper left hand corner of the plot.

The reduction in confounded information with increasing cost is a general trend across the population. Considering only those neurons sensitive to both cues in the stimulus set (defined as having a relative $rMI_X \geq 10\%$ for each cue, n=105), the median confounded information at zero cost is 23%, while the median confounded information at peak cost is 16% (the two are different at $p<1e-5$, rank sum). Since rate is a unidimensional measure, using a rate code to represent more than one cue necessarily leads to a confounded representation of the encoded quantities. The extra dimensions of spike timing allow a more independent representation of the localization cues, and in theory allow the cues to be decoded more independently, as well.

Coding trajectories for the entire population are summarized in Fig. 5.13. These plots show vectors like the arrow in Fig. 5.12D that point from the zero-cost position in the relative-information plot to the cutoff cost position (where the SDM MI has decayed to half of its peak value). The arrows are translated so the zero-cost position is at the origin; thus the lines represent the changes in relative information as cost increases. Only neurons for which both individual-cue MI_{peak} values are greater than 0.2 bits are considered. The gray lines represent the change in relative information from zero cost to cutoff cost for individual neurons, and the mean trajectory for the whole population is given as the thick black line.

Figure 5.13A shows the coding trajectories of neurons recorded with the SN/ILD stimulus set. Trajectories pointing to negative values on the abscissa represent cases where the percentage of MI_{FULL} devoted to coding ILD decreases over its zero cost value

when spike timing is taken into account; trajectories pointing upward represent cases in which the percentage devoted to SN increases with cost. There is remarkable consistency in the coding trajectories across the population, with the majority of trajectories heading in roughly the same direction. As the absolute timing of spikes becomes relevant in the code, the representation of SN increases, and the representation of ILD decreases. To test the significance of this result, the calculation of the population mean was repeated 1000 times with bootstrap sampling from the trajectory vectors. The results are shown in the inset, where the red dots correspond to the endpoints of the mean trajectory vectors from the bootstrapped data sets. Essentially, the cloud of red dots represents the two-dimensional confidence interval of the mean trajectory vector endpoint. All of the bootstrapped values lie in the second quadrant, indicating that the change in coding representation from ILD to SN with increasing temporal resolution is a general, significant trend of the population.

The behavior of the ITD/ILD and ABI/ILD coding trajectories are shown in Figs. 5.13B and 5.13C, respectively. For these sets, the general behavior is similar. However, the vectors are shorter because of the relatively small amount of MI revealed by the SMD analysis for both ITD and ABI. The trend with the ITD/ILD set is for MI_{ILD}/MI_{FULL} to increase at the expense of MI_{ITD}/MI_{FULL} and for the ABI/ILD set, MI_{ABI}/MI_{FULL} increases at the expense of MI_{ILD}/MI_{FULL}.

5.4 Discussion

5.4.1 Temporal representation of sound localization cues in IC

The main question addressed in this chapter is whether the temporal patterns of spike trains can enhance the representation of sound localization cues in the IC. Conceptually, such temporal information could be stimulus locked, e.g., by phase-locking to the waveform of the stimulus, or not stimulus locked. In the latter case, sound localization cues would be represented by changes in the temporal patterns of spiking that are not directly related to the stimulus' temporal waveform, as in Optican and Richmond's work in visual cortex (Optican and Richmond, 1987; Richmond and Optican, 1990). Of course most auditory neurons, including those in the IC, lock strongly to the stimulus envelope (Joris, 2003; Louage *et al.*, 2004). Thus evaluation of non-stimulus-locked temporal coding must control for these envelope responses; here we used random noise, for which envelope locking should not provide consistent information from one stimulus to the next.

An SDM analysis was used to look at temporal coding. An important check on this analysis is the fact that the zero-cost MI is the same as the discharge rate MI ($rMI \approx MI_0$ in Fig. 5.4C). Since the SDM method is based on stimulus parameter estimation, it provides a lower bound to the information available in the spike trains. For the zero cost case, the SDM method recovers all of the information available; however, this is not true at higher costs, since we know the method is insensitive to first spike latency information (discussed below in section 5.4.2). Thus, the information increment analysis (Fig. 5.5) should be looked upon as a lower bound to the extra temporal information that is available in spike trains.

The results show that encoding in spike timing potentially enhances the amount of information carried about localization cues in IC (Figs. 5.4-5.6). Significant timing-dependent increments were seen for all the cues except ITD, with the largest effects for SN. The lack of ITD-related temporal information suggests that ITD is represented by discharge rate alone in IC. This is consistent with the work of Carney and Yin (1989), who investigated the effects of ITD manipulation in a population of low-frequency IC neurons. Their raster plots of neurons responding to broadband noise at various ITDs (*cf.* their Figs. 10 and 11) show that there is little change in the timing of spikes to changes in ITD, rather, there is a large ITD-dependent gain change.

The position of the peak in MI versus cost functions (e.g., Fig. 5.2C-E) provides an estimate of the time scale at which spike timing provides the most information. The data in Fig. 5.7 show that localization cues in the IC are best decoded at a cost of ~100 s^{-1}, suggesting that the resolution of localization-related spike timing patterns in the IC is approximately 10 ms.

5.4.2 The nature of the temporal representation

When random noise was used to eliminate stimulus-waveform cues, the only temporal information remaining in IC neurons was that encoded in first-spike latency. Sound location has been shown to modulate the first-spike latency in both IC and auditory cortex (Brugge *et al.*, 1996; Furukawa and Middlebrooks, 2002; Sterbing *et al.*, 2003; Mrsic-Flogel *et al.*, 2005). While latency differences contribute to the distances measured with the SDM, in practice the variation in spike train distances caused by

latency are too small to have much effect on stimulus grouping, unless the analysis is confined to the first few spikes. The analysis of temporal information presented here does not address the role that first-spike latency may play in encoding sound localization cues; that will be discussed in the next chapter.

The frozen/random noise analysis (Figs. 5.7-5.9) shows that the temporal patterns are mainly locked to temporal features of the stimulus waveform, independent of static localization cue values. The largest stimulus-waveform effects are related to SN cues. Presumably these represent phase-locking to the temporal envelope of the stimulus induced by the sharp antiresonances in the SN stimuli. The strongest temporal information about SN occurs at BFs below the physiological range of SN cues in cats (Musicant *et al.*, 1990; Rice *et al.*, 1992). This suggests that the temporal increments for SN stimuli do not represent a specialization for representing SN. Instead, the temporal information is induced by spectral irregularities or temporal envelopes in general, as for example in speech (Bandyopadhyay and Young, 2004).

Given that the temporal information is locked to the stimulus, what is it about the stimulus that induces temporal information? In Fig. 5.14A, the first 30 ms of the temporal waveforms of two SN stimuli presented to a 2 kHz neuron are plotted. As can be seen from the plot, the differences in the temporal waveforms of these stimuli are quite subtle, even though the SN timing gain at 2 kHz was very high, in general (Fig. 5.6). Of course, in reality the band pass receptive field of the neuron will emphasize the stimulus differences around BF. Fig. 5.14C shows the same temporal waveforms, after filtering through a filter mimicking a linear auditory nerve fiber centered at 2 kHz (based on the model of Carney, 1993). When focused on the relevant frequency range, the temporal

waveforms are quite different. IC neurons are able to differentiate rather subtle differences in spectral notch based on the different temporal patterns induced by peripheral filtering.

Figures 5.14B and 5.14D show the same effect for stimuli centered on 11 kHz, where the SN timing gains are quite small. There are still differences between the filtered stimulus waveforms in this frequency range. The lack of extra information in spike timing in this frequency range is likely due to the better rate representation of these SN stimuli, rather than a decrease in the envelope differences.

Can the temporal information identified here be used by the auditory system for sound localization? In order to do so, there would have to be a template for the spike trains expected from a known stimulus, i.e. the stimulus would have to be recognized by the auditory system on the basis of its other properties. Then its location could be determined in part through the envelope induced by SN cues as demonstrated here. However, this source of information would be vulnerable to echoes and other environmental phase distortions, limiting its usefulness as an absolute localization cue.

A situation in which temporal cues might contribute is when comparisons of two stimuli occurring in the same acoustic environment are possible, e.g. in determining when a given sound source has changed location. In binaural-masking-level-difference experiments, random noises are more effective at masking interaural correlation differences than frozen noises (Breebaart and Kohlrausch, 2001), due to the uncertainty in the masker's interaural correlation. The results of this chapter suggest that minimum audible angles for random noise stimuli should be higher than for frozen stimuli, because of better encoding of SN cues in the latter case. Another situation where temporal

locking would be useful would be in the comparison of spike times across different neurons. This type of population encoding is not considered in this analysis.

Perhaps more surprising than the temporally-locked SN information is the temporally-locked ILD information. ILD is a static cue, yet its representation in the neural response benefits from spike times locked to the stimulus. When the ILD is changed, events in the stimulus that were sub-threshold could become super-threshold and cause the neuron to spike. These spikes would force the neuron into its refractory period, and may affect the position of the next burst of spikes. Thus, changes in ILD could cause a rearrangement in the peri-stimulus time histogram (PSTH). Of course, the same argument could be made for changes in ITD, which do not result in rearrangements in the PSTH; the mechanisms behind ILD and ITD encoding need to be further explored.

5.4.3 Differences among IC neuron classes

In the last chapter, the information about localization cues provided by IC neurons of three different response types was compared. That analysis, based only on discharge rate, found that although there were some differences among the neuron types, generally there was substantial overlap in the information provided by the three classes of neurons. The largest differences were for the type-V neurons, which provided information mainly about ITD and ABI. Type-V neurons stand out in the present analysis by showing larger MI increments than the other neuron types when spike timing is considered. Because type-V neurons are found only at low BFs and because most of the neurons in the low-BF sample were type V, it is not clear if the difference has to do with BF or with the

particular circuitry connected to the type-V neurons. An argument for the former is that the changes in temporal envelope produced by shifting the location of a spectral notch will be at higher envelope frequencies for high-BF neurons, compared to low-BF neurons. Given that neurons in the IC have a cutoff frequency in their modulation transfer functions of ~100 Hz (Langner and Schreiner, 1988; Krishna and Semple, 2000), it may be that the temporally coded information produced by changes in SN frequency is outside the neurons' modulation response regions for high-BF neurons.

5.4.4 The representation of multiple cues

These results show that spike-timing codes can reduce the confounded information in the response, allowing individual cues to be represented more independently (Fig. 5.12). Furthermore, the representation of cues in the response changes as a function of the decoding time resolution in a consistent manner, as illustrated by Fig. 5.13. ITDs are coded on the longest time scales, by spike rate. ILDs are coded at intermediate time scales, as increases in the SDM cost cause an increase in ILD information relative to ITD information, but a decrease relative to SN information. SN information is available at the shortest time scales, especially in low frequency neurons. Taken together, these results imply that spike timing could play an important role in multiplexing information onto spike trains, given appropriate decoding mechanisms.

6. Sound localization information in first spike latency

It is well known that several parameters of a stimulus, such as stimulus intensity in the auditory system or contrast in the visual system, can modulate the timing of the first spike in sensory neurons (Møller, 1975; Gawne *et al.*, 1996). In fact, the first spike latency has been shown to carry information in several sensory modalities, including the auditory (Furukawa and Middlebrooks, 2002), visual (Reich *et al.*, 2001), and somatosensory (Panzeri *et al.*, 2001) systems. As was shown in the last chapter, sound localization cues can also alter the first spike latency of single neurons, even when the noise carrier is changed on every presentation (Fig. 5.8). In this chapter, the role of first spike latency in the coding of sound localization cues in the IC is investigated.

The chapter will proceed as follows. In section 6.1, first spike latency coding will be motivated by a discussion of how first spikes differ from other spikes. In section 6.2, the first spike latency coding of both single and multiple localization cues is investigated, and in section 6.3, that information is compared with information conveyed by rate. The feasibility of general first spike latency codes is explored in section 6.4, and section 6.5 will summarize the major conclusions of this chapter.

6.1 First spikes are special

As was shown in Fig. 5.8 for the frozen/random noise studies, the timing of the first spike is unique in that it is the only spike whose timing depends more on the static parameters of the sound than on the carrier. Even within frozen noise responses, however, the timing of the first burst of spikes stands out as unique from the other bursts. Figure 6.1A shows the spike rasters recorded from a 1.3 kHz type-V neuron in response to the ITD/ILD stimulus set. The first burst of the response shows a clear dependence on ILD, with longer latencies occurring for more contralateral ILDs. Notice that the latency dependence on ILD is not a simple delay of the entire spike train, as the spike bursts across latency become more closely aligned as the stimulus progresses. The second burst of spikes has less latency dispersion than the first, the third has less than the second, until, about 25 ms after stimulus onset, the spike bursts align.

Once the latency dispersion has adapted away, the bursts remain aligned throughout the stimulus duration. This is emphasized in Fig. 6.1B, where a full 150 ms of the response is shown. Here, the spike trains of Fig. 6.1A have been shifted to align their first bursts with the first burst of the stimulus shown on the bottom, where the amount of delay was calculated as the time of the peak cross correlation between the first 12 ms of the PSTHs. Now that the start of the spike bursts line up, the responses slowly become unaligned, and remain so throughout the response.

Presumably, what is happening is that before stimulus onset, the neuron is well below threshold. When the stimulus turns on, the ILD determines a rate of depolarization, and

causes a latency dispersion accordingly. In the sustained portion of the stimulus, however, the neuron is never very far from threshold, and ILD-related differences in depolarization rate do not lead to much dispersion in the timing of spike bursts. This same effect was seen in cochlear nucleus neurons in response to pseudorandom noise of varying intensity (Møller, 1975).

Whatever the mechanism, the result is clear: first spikes are different. The latency variation of the first spike can potentially carry information unique to the stimulus. In the next few sections, this information will be quantified, related to the different localization cues, and compared with information in other codes.

6.2 First spike latency coding of localization cues

6.2.1 The coding of individual cues

In section 3.7 of the MI chapter, two methods were discussed for computing the information in first spike latency, MI_{FSL}. Neither method is ideal. The binless method tends to be less biased than the KDE method, but suffers from larger variance, and is not bound to return non-negative values. In most of the simulations (Fig. 3.14), the binless method outperforms the KDE method in terms of average error for the numbers of stimulus repetitions typically collected in these experiments. However, as discussed in section 3.7.3, one cannot make exact estimates of relative MI with the binless method, because the conditions of the cue-interaction equation (Eqn. 3.13) are not satisfied (section 3.8).

In Fig. 6.2, the MI_{FSL} estimated with the KDE method is plotted as a function of the MI_{FSL} estimated with the binless method for all of the neurons in the population. This analysis includes all of the first spikes that fall within a window ranging from 5 to 50 ms. No neurons were observed to have a first spike latency of less than 5 ms. MI estimates from the two methods are in good agreement (r=.95), and the difference between the two estimates decreases as more data is collected, as shown in the inset. Neither the binless estimates nor the KDE estimates of MI shown in this figure have been debiased. For the binless case, simulations show the bias to be fairly small, and debiasing would only serve to increase the variance. For the KDE case, the calculations are too computationally expensive to debias.

In this chapter, the binless method is used to compute the MI_{FSL}, unless otherwise stated. The KDE method will be used to compute the relative MI for comparing the coding of different cues, only. As in the last chapter, the analysis is restricted to neurons from which at least 20 repetitions of each stimulus were collected.

Figure 6.3 shows the MI_{FSL} for the full stimulus set as a function of BF. There were no differences in the information values across the stimulus sets, so they are not plotted separately. On average, first spike latency carries 0.65 bits of information about the full set of localization cues. Recall from Eqn. 3.31 that the first spike latency information can be broken down into two components, MI_{count} and MI_{timing}. MI_{count}, which denotes the information conveyed by the probabilities of getting at least one spike in the analysis window, is shown plotted as a function of BF in Fig. 6.3B. On average, MI_{count} accounts for only 0.18 bits of the first spike latency information. MI_{timing} is shown in Fig. 6.3C. When each neuron's MI_{timing} is multiplied by the probability of getting at least one spike

in the analysis window, the average information available in the timing of the first spike is 0.47 bits. Two neurons in Fig. 6.3C had negative MI_{timing} values, and were removed from the plot. These corresponded to instances where the probability of getting a spike in the analysis window was very low, and so the estimated information was off due to sampling issues.

As with the timing information uncovered by the SDM encoder, first spike latency information also differentially encodes the different localization cues. MI_{FSL} is plotted as a function of BF for each of the localization cues in Fig. 6.4. At least for the range of cue values tested in this study, first spike latency carries considerably more information about ILD than about either SN or ITD ($p_{ILD/SN}$<1e-10, $p_{ILD/ITD}$<1e-11, rank sum). The mean first spike latency information about ILD is not quite significantly different from ABI when corrected for multiple comparisons ($p_{ILD/ABI}$=.03, rank sum).

6.2.2 The coding of multiple cues

To look at the coding of multiple cues, relative MI plots are given in Fig. 6.5. The KDE method was used to make information estimates for these calculations. The points in these plots do not tend to lie on the diagonal, signaling the presence of confounded information. There is a large amount of confounded information in the first spike latency code, with a mean of 26%. This is significantly larger than the confounded information in a rate code over the same analysis window, which has a mean of only 15% (p<1e-13, rank sum). In comparison with the relative information plot for a rate code (Fig. 4.9B), the data for the ITD/ILD stimulus set tend to have less relative information about ITD.

This is consistent with the data of Fig. 6.4A, showing little overall information about ITD in first spike latency. Finally, the points in Fig. 6.5C tend to cluster toward the center of the plot (compare with the rate code data of Fig. 4.11C), indicating that FSL information tends to be equally affected by both level differences and overall level.

6.2.3 Frozen versus random noise

As noted in chapter 5, the SDM decoder was unable to recover any extra information in spike timing when random noise was used instead of frozen noise. Using the binless method, the information in first spike latency was estimated directly for the random noise cases. Figure 6.6 shows the first spike latency MI_{ILD} for random noise stimuli plotted as a function of the first spike latency MI_{ILD} for frozen noise. Most, but not all, of the information available in the frozen noise stimulus set is also available in the random noise set. In auditory nerve fibers, the first spike timing is largely determined by the acceleration of the peak pressure of the acoustic waveform (Heil and Irvine, 1997). The decrease in MI for the random noise set is likely due to the extra variability in the acceleration of peak pressure for the random noise stimuli.

Unfortunately, a comparison of MI_{SN} for the frozen/random noise stimulus set cannot be done, as none of the neurons measured with this stimulus set carries much information about SN. For frozen noise, the mean (± 1SD) of the MI_{SN} is 0.1 ± .09 bits, and for random noise it is 0.05 ± .07 bits.

6.3 Redundancy and independence of first spike latency codes

Of course, the obvious question to ask about the first spike latency code is how much *extra* information it provides. Does first spike latency tell us something different than rate, or can the information that first spike yields about the stimulus be deduced by looking at the spike rates?

First spike latency and rate cannot be completely independent. One of the components of first spike latency information is a contribution from MI_{count}, representing the information contained in whether or not there was a spike elicited by the stimulus within the analysis window, and MI_{count} is completely redundant with the information conveyed by spike rate. Even beyond this redundancy, however, first spike latency and rate tend to be correlated. Figure 6.7 shows the first spike times for one neuron plotted as a function of spike rate. As the number of spikes in the response increases, the latency of the first spike decreases.

To quantify exactly how much information is redundant or independent between rate and first spike latency, it is necessary to quantify how much information is available in their joint code, where the response is defined to be both rate and first spike latency. This computation is relatively straightforward. Using the chain rule for information (Cover and Thomas, 1991), the information between the stimuli and the joint code, MI(R,FSL;S), can be expressed as:

$$MI(R,FSL;S) = MI(R;S) + MI(FSL;S|R)$$

$$MI_{JOINT} = MI(R;S) + \sum_{r \in R} p(r) MI(FSL;S|R=r). \quad (6.1)$$

The first term on the right hand side is the rate information discussed in Chapter 4, and the second term represents the independent contribution of first spike latency to the information already available in rate.

Computing the joint information, while straightforward, is complicated by sampling problems. To carry out the computation in Eqn. 6.1, the information between first spike latency and the stimuli must be computed for each of the subsets of spike trains that contain a certain spike count. To help avoid sampling issues, the analysis in this section is limited to the neurons that were held through at least 100 stimulus repetitions (n=54). The binless method was used to estimate MI_{FSL}, as it tends to have less bias.

Figure 6.8A shows the information in the joint rate/FSL code plotted as a function of the information available in rate alone. The vertical offset of these points from the diagonal represents the extra information available by considering first spike timing; these values are given in the inset histogram. On average, first spike latency conveys about 0.3 bits of information about these stimuli over that conveyed by spike rate. Stated another way, approximately 54% of the information provided by first spike latency is independent of that provided by rate.

Figure 6.8B shows the information in the joint rate/FSL code, this time plotted as a function of the information available through first spike latency alone, with the inset giving the distribution of the extra information contributed by spike rate over that contributed by first spike latency. On average, spike rate conveys an extra 0.3 bits of information over that available in first spike latency alone. Roughly 56% of the rate information is independent of the first spike latency information. (Note that, within this 45 ms wide analysis window, the average rate MI is 0.65 bits).

Figure 6.9A shows MI_{JOINT} plotted as a function of the sum of the MIs conveyed by the individual codes. Recall from Eqn. 3.15 in section 3.3.2 of the MI chapter that the vertical offset of these points from the diagonal represents the interaction of synergy and redundancy within the coding variables. Apparently, the redundancy between rate and first spike latency outweighs the synergy, as the information in the joint code tends to be less than the combined MIs of the individual codes. This point is emphasized in Fig. 6.9B, where the synergy, MI(FSL;R|S), is plotted as a function of the redundancy, MI(FSL;R). The redundancy almost always outweighs the synergy.

6.4 Feasibility of first spike latency codes

Most studies measuring first spike latency information assume that the brain has an independent reference for stimulus onset from which to extract latency, and this has been the assumption made so far in this chapter. Critics of first spike latency codes argue that if latency is measured relative to a stimulus onset estimated from the population, the information conveyed by the first spike latency of single neurons will at the least decrease, if not disappear entirely, because first spike latency should be correlated across neurons. In this section, the MI conveyed by first spike latencies measured relative to stimulus onset is compared to the MI conveyed by first spike latencies measured relative to a realistic model of stimulus onset detection. Data from the SN/ILD stimulus set is excluded in this analysis, as those stimuli were resampled to shift the SN over the BF of each neuron, and so the stimuli are not the same across the population.

6.4.1 Similarity of FSL functions across the population

If all neurons in the population had similar first spike latency variation in response to different stimulus parameters, one might expect that measuring first spike latencies relative to a population-based stimulus onset cue would cause a reduction in the latency variation, decreasing the total available information. In this section, an algorithm is developed to detect the excitatory response latency of single neurons in the presence of spontaneous discharges. This algorithm is used to compute first spike latency functions for all of the neurons, such that their similarity can be directly measured.

To detect the latency of a single neuron, a binless method of first spike latency detection was employed in which spikes from all N repetitions of a stimulus were collapsed into a single response. For a neuron firing at spontaneous rate λ, the probability of getting n spikes in a time window t is well described by a Poisson process:

$$P_t(n) = \frac{(N\lambda t)^n e^{-N\lambda t}}{n!}. \tag{6.2}$$

The algorithm proceeds as follows. For every spike, the probability $P_{t_n}(\geq n)$ of observing a response at least as strong as the last n spikes actually observed in the interval t_n is computed under the assumption that the neuron is a Poisson process firing at its spontaneous rate:

$$P_{t_n}(\geq n) = 1 - \sum_{m=0}^{n-1} \frac{(N\lambda t_n)^m e^{-N\lambda t_n}}{m!}. \tag{6.3}$$

The spontaneous rate, λ, was computed as the average of the rates observed in the last 100 ms of the silent interval between stimuli.

The probability $P_{tn}(\geq n)$ is computed for n varying from the last 5 observed spikes to all of the spikes observed from stimulus onset to the current spike time. The first spike for which any of those probabilities exceeds a threshold of 10^{-6} is marked as the latency of the neuron to that stimulus. If this threshold is not exceeded within 30 ms of stimulus onset, the latency is left undefined. The inclusion of multiple windows is important for this algorithm to work. Long windows allow the algorithm to be sensitive to slow changes in the firing rate; short windows allow the algorithm to be accurate in the presence of fast firing rate changes. Figure 6.10 provides a schematic example of the algorithm.

First spike latency functions were estimated from all neurons recorded with the ITD/ILD and ABI/ILD stimulus sets. Figure 6.11 shows several examples of common first spike latency functions measured in response to the ITD/ILD stimulus set. It is clear that there is considerable variation in the first spike latency functions for different neurons, where some neurons show decreases in latency with increased ILD (Fig. 6.11A), some show increases in latency with increased ILD (Fig. 6.11B), some neurons show primarily binary responses (burst or no burst, depending on ILD; Fig. 6.11C), and some neurons show no variation in latency with ILD (Fig. 6.11D). ITD does not tend to affect the first spike latency functions, as expected from the MI results of Fig. 6.4.

To test the variation in first spike latency functions across the population, cross-correlation coefficients were computed for the first spike latency functions of every pair of neurons recorded with the same stimulus set. Recall that the first spike latency was left undefined for cases where the firing rate did not differ significantly from the spontaneous rate over the first 30 ms. In computing the correlation coefficients, the

analysis was restricted to pairs of neurons that shared defined first spike latencies for at least 15 common stimuli. Histograms of these correlation coefficients are presented in Fig. 6.12.

On average, the first spike latency functions of neurons in response to the ABI/ILD stimulus set (Fig. 6.12A) are weakly correlated, with a mean correlation coefficient of 0.30. There is much more variation in response to the ITD/ILD stimulus set, with a number of neurons having uncorrelated or even anti-correlated responses (Fig. 6.12B), though the mean is still slightly positive at 0.14. The differences between the pair-wise cross-correlation coefficients to the two stimulus sets probably stems from two sources. First, almost all of the neurons in the population show a decrease in latency with increasing ABI, whereas ITD does not tend to affect latency. Also, the increases in latency with increased ILD, which occurs mainly in neurons that are excited by localization cues representing ipsilateral sources, occurs more often in neurons with low BFs. There are more neurons with low BFs in the population to which ITD/ILD stimuli were presented than ABI/ILD stimuli (see Fig. 6.4).

6.4.2 A coincidence detector model of onset detection

To estimate stimulus onset, all neurons recorded with the same stimulus set are lumped into a pseudo-population, as if they had been recorded simultaneously in the same cat. They are then used as input to a coincidence detector that will signal stimulus onset when the neural activity across the population reaches a certain threshold. In the

auditory system, such a model could mimic the activity of onset neurons in the ventral cochlear nucleus (Winter and Palmer, 1995; Kalluri and Delgutte, 2003).

Figure 6.13A provides a schematic of the algorithm. To estimate the population onset time for each stimulus, one spike train was selected at random from each neuron in the population, from the set of spike trains recorded from that neuron in response to that stimulus. Every spike in these trains was then convolved with an exponential tail (with a user-defined decay time constant, here 1 ms) to mimic the excitatory post-synaptic potentials (EPSPs) that these spikes would induce on the coincidence detector neuron (Fig. 6.13B). All of these trains were summed together, with equal weight, as a rough estimate of the total EPSP trace that a coincidence detector neuron might experience. The results of one such run are shown in Fig. 6.13C, giving the summed EPSP as a function of time from stimulus onset. The temporal resolution of the EPSP trace is 0.1 ms.

The summed EPSP trace from a single run is quite noisy, because the activity across the neurons is typically sparse. To smooth out the noise, the above procedure was repeated 500 times, with a different random set of spike trains sampled from the input neurons each time. Sampling was done with replacement, in bootstrap fashion, and the resultant post-synaptic potential traces were averaged together. Altogether, this process mimics sampling from a much larger population of neurons than were recorded.

Figure 6.13D shows the average summed EPSP trace of the coincidence detector for three stimuli from the ITD/ILD stimulus set. All three traces show sharp deflections at well-defined times after stimulus onset. Threshold is defined as halfway between the mean spontaneous activity (defined as the average activity over the last 100 ms of the

silent interval between stimuli) and the mean driven activity (defined as the activity in a window between 50 and 300 ms after stimulus onset), where the mean is taken over all 25 stimuli in the set. When the average summed EPSP trace crosses threshold, the coincidence detector signals stimulus onset.

Figure 6.14 shows the population onset time estimates from the coincidence detector as a function of stimulus for the ITD/ILD and ABI/ILD stimulus sets. Results are shown for three different time constants of exponential decay. There is considerable variation in the estimate of onset time with stimulus identity, suggesting that first spike latency information measures assuming an invariant stimulus onset reference may not be correct. There are also considerable differences across the two stimulus sets, with the ABI/ILD stimuli inducing much more latency variation than the ITD/ILD stimuli. This is probably because the ABI/ILD stimuli span a larger range of sound levels than the ITD/ILD stimuli. An exponential decay constant of 1 ms was chosen for the analysis below.

6.4.3 First spike latency information using a population onset reference

The MI_{FSL} was re-computed using the stimulus onset time derived from the coincidence detector as the onset reference. To do this, spike times for each stimulus were measured relative to the population onset, where spikes occurring before the population onset signal were discarded. This procedure is outlined in Fig. 6.15.

The rasters of one neuron's response to the ITD/ILD stimulus set are shown in Fig. 6.15A, as a function of time from stimulus onset. The first spike in each line (occurring in the window [5-50] ms) is shown as a dark blue dot, non-first spikes are shown as thin

green dots. This neuron clearly shows first spike latency modulation to ILD. The population-derived stimulus onset time is shown as a red line; in this case, the neuron's own first spike latency function correlates well with the population-derived stimulus onset.

Figure 6.15B shows the same rasters, plotted as a function of the population-derived stimulus onset time. This time, the first spikes after the population-onset (occurring in the window [0-45] ms) are shown in dark blue, with all of the other spikes in green. The population-derived stimulus onset time is now a flat line at zero.

If an individual neuron's first spike latency function is correlated with the stimulus onset time estimated from the population, measuring the latency of first-spike times relative to the population-onset time will bring the spike bursts closer together. This will in turn make the stimuli less distinguishable and decrease the information conveyed by first spike latency about stimulus identity. This is the case for the neuron of Figs. 6.15A and B. The full MI_{FSL} assuming an independent zero-onset time reference is 0.91 bits, based on the dark blue spikes from Fig. 6.15A. For comparison, the full MI_{FSL} assuming a population-derived onset reference, based on the blue spikes of Fig. 6.15B, is only 0.55 bits.

Neurons whose first spike latency functions are anti-correlated with the population onset response will have spike bursts that overlap less when measured relative to the population onset time, making the stimuli more distinguishable and increasing the information conveyed by first spike latency. This behavior is illustrated for one example neuron in Fig. 6.15C and D. For this neuron, measuring the spike times relative to the population-derived stimulus onset *induces* dispersion in the spike bursts, and causes the

MI_{FSL} to increase from its zero-onset reference value of 1.02 bits to its population-onset reference value of 1.37 bits.

The correlation of a neuron's first spike latency function with the population onset is not the only factor determining whether information will increase or decrease. Usually, the latency variation of the first burst of spikes is different from that of subsequent bursts, as discussed in section 6.1 and illustrated in Fig. 6.1. Since the neurons do not necessarily fire a spike in the first burst, the next few bursts can also play a role in first spike latency information transfer. Thus, even when the first bursts of spikes overlap more when corrected for the population estimate of stimulus onset, reducing information, the second bursts often overlap less, increasing the information. This phenomenon explains why the MI_{FSL} of the neuron in Fig. 6.15A and B did not show a more dramatic increase when spikes were measured relative to the population-onset.

Across the population of neurons, the competing processes illustrated in Fig. 6.15 work against each other. In Fig. 6.16, the MI_{FSL} computed using first spikes measured relative to the population-derived onset are plotted as a function of the MI_{FSL} measured assuming an independent, zero onset time reference. On average, there is a small net *increase* in information when spike times measured relative to a population-derived stimulus onset, though this is not statistically significant (mean, independent zero-time reference: 0.7 ± 0.4 bits; mean, population reference: 0.8 ± 0.4 bits; p=0.26, ranksum).

6.5 Discussion of first spike latency results

6.5.1 First spike latency information by cue type

A number of studies have shown that that the spatial location of a sound can modulate the first spike latency of auditory neurons, and indeed one can even build spatial receptive fields out of first spike latencies alone (Brugge *et al.*, 1996; Eggermont, 1998; Sterbing *et al.*, 2003; Behrend *et al.*, 2004). There are few studies, however, addressing exactly how the first spike latency relates to spatial location, and on how the first spike latency representation might differ from other representations, such as spike rate. The focus of this chapter is on how first spike latency might be used as a neural code for the various localization cues.

First spike latency differentially encodes sound localization cues of different types (Fig. 6.4). ITD information is poorly represented by first-spike latency, which furthers the claim of the last chapter that ITD information is carried by a spike rate code, only. With an SDM decoder, ILD and ABI cues were found to have a modest spike-timing related information gain, while SN cues should the largest gain (Figs. 5.5-5.6). With first spike latency coding, the situation is reversed; ILD and ABI cues are coded most strongly, while SN cues are modestly represented.

6.5.2 Information in multiple codes

A joint rate/first spike latency code contains more information than either code alone, indicating some degree of independence in the representation (Fig. 6.8). However, the representations are not completely independent (Fig 6.7), and the redundancy between

rate and first spike latency overwhelms any synergistic coding effects they might have (Fig. 6.9).

It has been suggested that a joint code consisting of spike count and first spike latency (Furukawa and Middlebrooks, 2002) or mean spike latency (Nelken *et al.*, 2005) is sufficient to capture the majority of information transferred by auditory neurons. For the analysis window used in this chapter, [5-50] ms, this also seems to be true. The average MI_{JOINT} for a rate/first spike latency code (considering only those neurons recorded for at least 100 repetitions) is 0.99 ± 0.40 bits. Meanwhile, the average peak MI from the SDM decoder on the same analysis window is almost identical at 0.97 ± 0.47 bits.

For longer analysis windows, the joint code still does remarkably well. Re-analyzing all of the data for spikes arriving anywhere from stimulus onset to 20 ms past stimulus offset yields an average MI_{JOINT} of 1.45 ± 0.47 bits, with the average SDM MI_{peak} at 1.63 ± 0.73 bits. However, the coding of the individual cues is quite different. Figure 6.17 shows MI_{JOINT} for a rate/first spike latency code plotted as a function of SDM MI_{peak} for the ILD and SN cues. While the joint code recovers all of the information about ILD, it systematically underestimates the information about SN, especially in low BF units where there is significant spike timing gain (Fig. 5.6).

6.5.3 First spike latency as a general coding mechanism

Of course, neural codes that depend upon an independent reference for stimulus onset are problematic, because in real life organisms must derive stimulus onset from the population of responsive sensory neurons. In the auditory system, onset cells, which

receive across-frequency input from hundreds of cells, are uniquely suited to this role, and modeling studies have shown that their response looks much like a classical coincidence detector (Kalluri and Delgutte, 2003). The results of section 6.4 provide evidence for the counter-intuitive effect that using a population reference of stimulus onset can in some cases increase the information transferred by a first spike latency code, and on average does not decrease the information. This is mainly due to the variation in first spike latency functions across the population. Neurons that do not show latency variation with respect to stimulus onset may show information when latency is measured relative to the population onset. That is, while the population latency measure degrades the information in some neurons, it induces information in other neurons. Thus, first spike latency codes are robust to different measures of onset time, and are a feasible, general mechanism of information transfer in sensory systems.

7. Conclusions

This chapter wraps up the findings discussed in the previous chapters by putting them into the broader context of general sensory processing mechanisms. It begins with a summary of the major findings (section 7.1), continues with a discussion of future directions (section 7.2), and concludes with a discussion of MI as a general technique for probing sensory representations (section 7.3).

7.1 Summary of the main conclusions

The main finding of this work is that there appear to be a diversity of coding mechanisms for the three different localization cues. At the level of the IC, ITD information is encoded by spike rate. The results of chapter 6 indicate that first spike latency codes ITD only weakly (Fig. 6.4), and in chapter 5, no extra information was recovered by the SDM decoder that was not present in rate (Fig. 5.5). ILD information is encoded by both first spike latency and spike rate, and in fact, the joint rate/FSL code appears to capture all of the information available (Fig. 6.16). The coding of SNs is

dependent on BF (Figs. 4.7 and 5.6). At high BFs, the majority of neurons code SN by spike rate and, to a lesser extent, by first spike latency. Low BF neurons, however, have little SN information in spike rate, and seem instead to represent SN by spike timing distributed throughout the duration of their response.

The coding of SN in low BF neurons is an interesting case. As noted in section 5.4.2, this information has little to do with sound localization, as SN cues do not exist in the low frequency region of cat HRTFs (Musicant *et al.*, 1990; Rice *et al.*, 1992); the response here has more to do with the coding of general spectral shapes. It is interesting to ponder how a static spectral shape might be deduced from a temporal code. The question is probably ill posed, however, and reflects a bias in the experimental design. In reality, the distinction between spectral events and temporal events becomes blurred in the low frequency region, and really one should consider the coding of combined spectro-temporal events.

The differential coding of the cues implies that information about different cues could be multiplexed onto the responses of single neurons, using different codes. The relative MI analyses of section 5.3 investigated this multiplexing at different temporal resolutions, finding that the percentage of information encoding each cue changes in a characteristic way with the temporal resolution of the decoder (Figs. 5.11 through 5.13). Adding a first spike latency code on top of the SDM decoder will only increase the independent representation of the three localization cues.

Using different coding schemes for different cues implies that downstream neurons may be able to decode these cues independently. To compute the position of a sound source, information from the three cues must eventually be integrated together.

However, the accuracy with which each cue can contribute to the overall location percept changes depending on sound conditions, such as the frequency content of the stimulus, temporal characteristics of the stimulus, or the reverberant qualities of the environment (Wightman and Kistler, 1992; MacPherson and Middlebrooks, 2002; Faller and Merimaa, 2004). If information about the individual cues is available in cortical areas, the relative weight with which each cue contributes to the location percept could be changed depending on sound conditions.

7.2 Future work

This section is devoted to interesting questions and data anomalies that have come up during the course of this research, but that were too off-topic to pursue.

7.2.1 Information carried by groups of neurons

One of the questions not addressed by this study is how information about sound localization cues is stored across the population of neurons. Population codes allow the added freedom of storing information in the relative timing of spikes across neurons, and so has potentially much greater information capacity. This extra flexibility also makes it much more difficult to study, however. Typically, the number of possible response bins grows exponentially with the number of neurons, requiring simultaneous recordings from groups of neurons for extraordinarily long periods of time.

7.2.2 Latency dispersion

One of the most interesting results observed in this work is the adaptation of the first spike latency dispersion discussed in section 6.1. If the latency dispersion does carry information that is used by the brain, then it would be useful to understand the process more fully. What are the properties of acoustic stimuli that cause this latency dispersion? How does the dispersion adapt away, and what properties of the stimulus cause it to come back again? Furthermore, is this process helpful in stimulus coding, or detrimental? It is interesting that the dispersion only occurs at the beginning of the stimulus. Could it somehow be related to stimulus grouping, or segregation?

7.2.3 ILD sensitivity and interaural correlation

In the initial studies using the random noise stimuli, an error was made where the stimulus set to each ear had a different noise carrier. Although all of the static localization cues were properly applied across the two ears, the interaural correlation was unintentionally removed. The interesting thing is that in a subset of neurons, the ILD sensitivity changed under this condition. Figure. 7.1 illustrates this effect for two example neurons. In each plot, the mean and the standard deviation of the spike counts are plotted as a function of stimulus. There are huge differences in the rate responses to the frozen, correlated noises and the random, uncorrelated noises. In Fig. 7.1A, the ILD sensitivity is much greater to the correlated noise condition, whereas in Fig. 7.1B, the ILD sensitivity is greater to uncorrelated noise. (Actually, the neuron of Fig. 7.1A is

probably more sensitive to overall level than to ILD; see the caption.) It seems the degree of interaural correlation plays a large role in the extent of the ILD sensitivity for IC neurons.

The neurons of Fig. 7.1 represent extreme examples. The responses of the other nine neurons for which this data were taken are shown in Fig. 7.2. The data are in the same format as Fig. 7.1, except now an asterisk has been placed below responses that are significantly different at either the 5% (black) or 1% (red) level, according to a Wilcoxson ranksum test. The majority of responses are significantly different. As a comparison, Fig. 7.3 shows the same analysis for neurons studied with the random noise stimuli presented correctly, i.e., correlated between the two ears but different on each repetition. Fourteen neurons were studied with this stimulus set; the responses of nine randomly chosen neurons are shown. In general, the responses of correlated frozen noise and correlated random noise are not significantly different from one another.

Unfortunately, no data were taken in the same neuron comparing random correlated noise to random uncorrelated noise, so it cannot be said for certain that the only difference is the interaural correlation. However, the degree to which ILD sensitivity is sensitive to interaural correlation is not well studied, and should be pursued. It is possible that studies of interaural correlation sensitivity may aid in distinguishing the types of inputs to each neuron, and may either correlate with the physiological cell type or form the basis for sub-categories within those classes.

7.3 Discussion of MI as a technique for exploring neural coding

One of the goals of this thesis is to evaluate MI as a technique for exploring questions of sensory processing. The results are mixed. There are three main difficulties with the application of MI to neural encoding:

1) When using information theory to characterize neural responses, there are always questions of interpretation. It is worthwhile to stress that MI is a measure of neural *sensitivity*, not *selectivity*; this study does not provide insight into how information is functionally distributed across physiological classes or the functional coding schemes actually employed. In other words, these results tell us if information is contained in the response, not whether (or how) it might be used behaviorally. For example, it is possible that the SN information conveyed by the type-I population is behaviorally distinct from the SN information conveyed by the type-O population. The tight tuning of the type-I neurons would make them ideal for conveying the spectral content of a sound, whereas the type-O pathway could be used for processing spectral information related to sound localization. This would agree with lesion studies of the DCN (Sutherland, 1998; May, 2000), which have been shown in cats to cause selective deficits in elevation localization, as type-O cells receive a direct input from the DCN (Davis, 2002). The type of analysis done here cannot distinguish these possibilities.

2) MI depends on the particular stimulus set employed, complicating its interpretation. For example, a neuron that only responds to one stimulus in a set (a grandmother cell, or in this context, a cell with extremely narrow tuning to sets of localization cues) could be a feature detector, and very important for stimulus coding. The MI in the response of such a neuron, however, would be small, and the more stimuli that are in the set, the smaller it would be. This is because the MI quantifies the

information carried about the entire stimulus set; if most stimuli do not cause a response, the neuron is usually uninformative.

3) Finally, the sheer amount of data required to make accurate estimates of MI is restrictive. The number of different tests and cases one can perform in a given experiment is limited by the number of repetitions one can collect, and the amount of data required typically grows exponentially as the number of stimuli increases. Thus, to make estimates of MI it is necessary to study only a small aspect of the neuron's response properties.

The results are not all negative, however. Most of the results of this study would not have been uncovered without the use of information theory. The differential coding of the localization cues described above was only evident because of the information theoretic approach. Furthermore, using MI to quantify neural sensitivity allows the identification of possible sources of information that do not take on a systematic form and might otherwise be overlooked. For example, the rate profiles in Figs. 4.5 and 4.10 show that type-I neurons provide a simpler, more intuitive representation of both SN and ILD that is more uniform across the population than type-O neurons. Nevertheless, there is little difference between the populations in terms of the information provided, and both populations may contribute to the representation of these cues. Stecker and Middlebrooks (2003) showed that a neural network model could extract spatial location from distributed codes in auditory cortex. Responses that are not systematic across a population of cells may still be useful to the brain in deciphering stimulus attributes; it is merely necessary that neurons change their response when source location changes.

Overall, MI is a useful technique for exploring sensory representations, but only in conjunction with other techniques. The MI experiments must be guided by studies that first fully explore the stimulus space and find a subspace of stimuli to investigate more deeply. Finally, some way of testing whether the information within a particular stream or code is actually used by the animal must be discovered. The feasibility study of first spike latency coding (section 6.4) is a step in this direction. Ultimately, though, direct tests of decoding mechanisms will probably need to be made in higher cortical areas.

8. References

Aertsen AM, Johannesma PI (1981) The spectro-temporal receptive field. A functional characteristic of auditory neurons. Biol Cybern 42:133-143.

Aitkin LM, Anderson DJ, Brugge JF (1970) Tonotopic organization and discharge characteristics of single neurons in nuclei of the lateral lemniscus in cat. J Neurophysiol 33:421-440.

Aitkin LM, Martin RL (1987) Neurons in the inferior colliculus of cats sensitive to sound-source elevation. Hear Res 50:97-105.

Aitkin LM, Martin RL (1990) The representation of stimulus azimuth by high best-frequency azimuth-selective neurons in the central nucleus of the inferior colliculus of the cat. J Neurophysiol 57:1185-1200.

Aitkin LM, Webster WR, Veale JL, Crosby DC (1975) Inferior colliculus. I. Comparison of response properties of neurons in central, pericentral, and external nuclei of adult cat. J Neurophysiol 38:1196-1207.

Bandyopadhyay S, Young ED (2004) Discrimination of voiced stop consonants based on auditory nerve discharges. J Neurosci 24:531-541.

Benevento LA, Coleman PD (1970) Responses of single cells in cat inferior colliculus to binaural click stimuli: combinations of intensity levels, time differences, and intensity differences. Brain Res 17:387-405.

Behrend O, Dickson B, Clarke E, Jin C, Carlile S (2004) Neural responses to free field and virtual acoustic stimulation in the inferior colliculus of the guinea pig. J Neurophysiol 92:3014-3029.

de Boer E, de Jongh HR (1978) On cochlear encoding: Potentialities and limitations of the reverse-correlation technique. J Acoust Soc Am 63:115-135.

Boudreau JC, Tsuchitani C (1968) Binaural interaction in the cat superior olive S segment. J Neurophysiol 31:442-454.

Brand A, Behrend O, Marquardt T, McAlpine D, Grothe B (2002) Precise inhibition is essential for microsecond interaural time difference coding. Nature 417:543-547.

Breebaart J, Kohlrausch A (2001) The influence of interaural stimulus uncertainty on binaural signal detection. J Acoust Soc Am 109:331-345.

Brugge JF, Reale RA, Hind JE (1996) The structure of spatial receptive fields of neurons in primary auditory cortex of the cat. J Neurosci 16:4420-4437.

Brunso-Bechtold JK, Thompson GC, Masterton RB (1981) HRP study of the organization of auditory afferents ascending to central nucleus of inferior colliculus in cat. J Comp Neurol 197:705-722.

Caird D, Klinke R (1987) Processing of interaural time and intensity differences in the cat inferior colliculus. Exp Brain Res 68:379-392.

Cant NB (1992) The cochlear nucleus: Neuronal types and their synaptic organization. In: The mammalian auditory pathway: Neuroanatomy (Webster DB, Popper AN, Fay RR, eds.), pp66-116. New York: Springer-Verlag.

Carney LH (1993) A model for the responses of low-frequency auditory-nerve fibers in cat. J Acoust Soc Am 93:401-417.

Carney LH, Yin TCT (1989) Responses of low-frequency cells in the inferior colliculus to interaural time differences of clicks: excitatory and inhibitory components. J Neurophysiol 62:144-161.

Chase SM, Young ED (2005) Limited segregation of different types of sound localization information among classes of units in the inferior colliculus. J Neurosci 25:7575-7585.

Chase SM, Young ED (2006) Spike-timing codes enhance the representation of multiple simultaneous sound-localization cues in the inferior colliculus. J Neurosci 26:3889-3898.

Chechik G, Globerson A, Tishby N, Anderson MJ, Young ED, Nelken I (2001) Group redundancy measures reveal redundancy reduction in the auditory pathway. In: Advances in Neural Information Processing Systems 14: Proceedings of the 2001 Neural Information Processing Systems Conference (Dietterich TG, Becker S, Ghahramani Z, eds.). Cambridge, MA: MIT Press.

Cover TM, Thomas JA (1991) Elements of information theory. New York: John-Wiley & Sons.

Davis KA (2002) Evidence of a functionally segregated pathway from dorsal cochlear nucleus to inferior colliculus. J Neurophysiol 87:1842-1835.

Davis KA, Ramachandran R, May B (1999) Single-unit responses in the inferior colliculus of decerebrate cats. II. Sensitivity to interaural level differences. J Neurophysiol 82:164-175.

Davis KA, Ramachandran R, May B (2003) Auditory processing of spectral cues for sound localization in the inferior colliculus. JARO 4:148-163.

Delgutte B, Joris PX, Litovski RY, Yin TCT (1995) Relative importance of different acoustic cues to the directional sensitivity of inferior-colliculus neurons. In: Advances in hearing research (Manley GA, Klump GM, Köppl C, Fastl H, Oeckinghaus H, eds.), pp288-299. River Edge, NJ: World Scientific.

Delgutte B, Joris PX, Litovski RY, Yin TCT (1999) Receptive fields and binaural interactions for virtual-space stimuli in the cat inferior colliculus. J Neurophysiol 81:2833-2851.

Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001) Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. J Neurophysiol 85:1220-1234.

Efron B, Tibshirani RJ (1998) Introduction to the bootstrap. Boca Rotan, FL: CRC Press LLC.

Eggermont JJ (1998) Azimuth coding in primary auditory cortex of the cat. II. Relative latency and interspike interval representation. J Neurophysiol 80:2151-2161.

Eggermont JJ, Johannesma PM, Aertsen AM (1983) Reverse-correlation methods in auditory research. Q Rev Biophys 16:341-414.

Faller C, Merimaa J (2004) Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. J Acoust Soc Am 116:3075-3089.

Furukawa S, Middlebrooks JC (2002) Cortical representation of auditory space: information-bearing features of spike patterns. J Neurophysiol 87:1749-1762.

Gawne TJ, Kjaer TW, Richmond, BJ (1996) Latency: Another potential code for feature binding in striate cortex. J Neurophysiol 76:1356-1360.

Geisler CD, Rhode WS, Hazelton DW (1969) Responses of inferior colliculus neurons in the cat to binaural acoustic stimuli having wide-band spectra. J Neurophysiol 32:960-974.

Glavinović MI (1996) Comparison of Parzen density and frequency histogram as estimators of probability density functions. Eur J Physiol 433:174-179.

Goldberg JM, Brown PB (1969) Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. J Neurophysiol 32:613-636.

Golomb D, Hertz J, Panzeri S, Treves A, Richmond B (1997) How well can we estimate the information carried in neuronal responses from limited samples? Neural Comput 9:649-665.

Guinan JJ Jr, Guinan SS, Norris BE (1972a) Single auditory units in the superior olivary complex. I: Responses to sounds and classifications based on physiological properties. Int J Neurosci 4:101-120.

Guinan JJ Jr, Norris BE, Guinan SS (1972b) Single auditory units in the superior olivary complex. II: location of unit categories and tonotopic organization. Int J Neurosci 4:147-166.

Heil P, Irvine DRF (1997) First-spike timing of auditory-nerve fibers and comparison with auditory cortex. J Neurophysiol 78:2438-2454.

Huang AY, May BJ (1996) Sound orientation behavior in cats. II. Mid-frequency spectral cues for sound localization. J Acoust Soc Am 100:1070-1080.

Imig TJ, Bibikov NG, Poirier P, Samson FK (2000) Directionality derived from pinna-cue spectral notches in cat dorsal cochlear nucleus. J Neurophysiol 83:907-925.

Irvine DR (1987) Interaural intensity differences in the cat: changes in sound pressure level at the two ears associated with azimuthal displacements in the frontal horizontal plane. Hear Res 26:267-286.

Irvine DRF (1992) Physiology of the auditory brainstem. In: The mammalian auditory pathway: Neurophysiology (Popper AN, Fay RR, eds.), pp153-231. New York: Springer-Verlag.

Irvine DRF, Gago G (1990) Binaural interaction in high-frequency neurons in inferior colliculus of the cat: effects of variations in sound pressure level on sensitivity to interaural intensity differences. J Neurophysiol 63:570-591.

Jeffress LA (1948) A place theory of sound localization. J Comp Physiol Psychol 41:35-39.

Jiang D, McAlpine D, Palmer AR (1997) Detectability index measures of binaural masking level difference across populations of inferior colliculus neurons. J Neurosci 17: 9331-9339.

Johnson DH, Gruner CM, Baggerly K, Seshagiri C (2001) Information-theoretic analysis of neural coding. J Comp Neurosci 10:47-69.

Joris, PX (2003) Interaural time sensitivity dominated by cochlea-induced envelope patterns. J Neurosci 23:6345-6350.

Joris PX, Carney LH, Smith PH, Yin TCT (1994) Enhancement of neural synchrony in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency. J Neurophysiol 71:1022-1036.

Joris PX, Smith PH (1998) Temporal and binaural properties in dorsal cochlear nucleus and its output tract. J Neurosci 18:10157-10170.

Joris PX, Yin TC (1995) Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences. J Neurophysiol 73:1043-1062.

Kalluri S, Delgutte B (2003) Mathematical models of cochlear nucleus onset neurons: I. Point neuron with many weak synaptic inputs. J Comput Neurosci 14:71-90.

Kiang NY, Sachs MB, Peake WT (1967) Shapes of tuning curves for single auditory–nerve fibers. J Acoust Soc Am 42:1341-1342.

Kjaer TW, Hertz JA, Richmond BJ (1994) Decoding cortical neuronal signals: network models, information estimation and spatial tuning. J Comput Neurosci 1:109-139.

Krishna BS, Semple MN (2000) Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. J Neurophysiol 84:255-273.

Kuhn GF (1977) Model for the interaural time differences in the azimuthal plane. J Acoust Soc Am 62:157-167.

Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. J Neurophysiol 60:1799-1822.

Liberman MC (1991) Central projections of auditory-nerve fibers of different spontaneous rate. I. Anteroventral cochlear nucleus. J Comp Neurol 313:240-258.

Licklider JCR (1948) The influence of interaural phase relations upon the masking of speech by white noise. J Acoust Soc Am 100:490-503.

Litovsky RY, Delgutte B (2002) Neural correlates of the precedence effect in the inferior colliculus: effect of localization cues. J Neurophysiol 87:976-994.

Louage DHG, van der Heijden M, Joris PX (2004) Temporal properties of responses to broadband noise in the auditory nerve. J Neurophysiol 91:2051-2065.

Louage DHG, van der Heijden M, Joris PX (2005) Enhanced temporal response properties of anteroventral cochlear nucleus neurons to broadband noise. J Neurosci 25:1560-1570.

Macpherson EA, Middlebrooks JC (2002) Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. J Acoust Soc Am 111:2219-2236.

MacKay DJC (2002) Information theory, inference, and learning algorithms. Cambridge University Press.

Macrae AW (1971) On calculating unbiased information measures. Psych Bull 75:270-277.

Manis PB, Marx SO (1991) Outward currents in isolated cochlear nucleus neurons. J Neurosci 11:1865-2880.

May BJ (2000) Role of the dorsal cochlear nucleus in the sound localization behavior of cats. Hear Res 148:74-87.

Merzenich MM, Reid MD (1974) Representation of the cochlea within the inferior colliculus of the cat. Brain Res 77:397-415.

Middlebrooks JC, Clock AE, Xu L, Green DM (1994) A panoramic code for sound location by cortical neurons. Science 264:842-844.

Middlebrooks JC, Green DM (1991) Sound localization by human listeners. Annu Rev Psychol 42:135-159.

Mills AW (1958) On the minimum audible angle. J Acoust Soc Am 30:237-246.

Møller AR (1975) Latency of unit responses in cochlear nucleus determined in two different ways. J Neurophysiol 38:812-821.

Moore GP, Perkel DH, Segundo JP (1966) Statistical analysis and functional interpretation of neuronal spike data. Annu. Rev. Physiol. 28:493-522.

Mrsic-Flogel TD, King AJ, Schnupp JWH (2005) Encoding of virtual acoustic space stimuli by neurons in ferret primary auditory cortex. J Neurophysiol 93:3489-3503.

Musicant AD, Chan JCK, Hind JE (1990) Direction-dependent spectral properties of cat external ear: New data and cross-species comparisons. J Acoust Soc Am 30:237-246.

Nelken I, Chechik G, Mrsic-Flogel TD, King AJ, Schnupp JWH (2005) Encoding stimulus information by spike numbers and mean response time in primary auditory cortex. J Comput Neurosci 19:199-221.

Nelken I, Kim PJ, Young ED (1997) Linear and nonlinear spectral integration in type IV neurons of the dorsal cochlear nucleus. II. Predicting responses with the use of nonlinear models. J Neurophysiol 78:800-811.

Nordeen KW, Killackey HP, Kitzes LM (1983) Ascending auditory projections to the inferior colliculus in the adult gerbil, Meriones unguiculatus. J Comp Neurol 214:131-143.

Oliver DL, Beckius GE, Bishop DC, Kuwada S (1997) Simultaneous anterograde labeling of axonal layers from lateral superior olive and dorsal cochlear nucleus in the inferior colliculus of cat. J Comp Neurol 382:215-229.

Oliver DL, Huerta MF (1992) Inferior and superior colliculi. In: The mammalian auditory pathway: Neuroanatomy (Webster DB, Popper AN, Fay RR, eds.), pp168-221. New York: Springer-Verlag.

Optican LM, Gawne TJ, Richmond BJ, Joseph PJ (1991) Unbiased measures of transmitted information and channel capacity from multivariate neuronal data. Biol. Cybern. 65:305-310.

Optican LM, Richmond BJ (1987) Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex I. Information theoretic analysis. J Neurophysiol 57:162-178.

Palmer AR, Jiang D, McAlpine D (2000) Neural responses in the inferior colliculus to binaural masking level differences created by inverting the noise in one ear. J Neurophysiol 84:844-852.

Paninski L (2003) Estimation of entropy and mutual information. Neural Comput 15:1191-1253.

Panzeri S, Petersen RS, Schultz SR, Lebedev M, Diamond ME (2001) The role of spike timing in the coding of stimulus location in rat somatosensory cortex. Neuron 29:769-777.

Panzeri S, Treves A (1996) Analytical estimates of limited sampling biases in different information measures. Network: Comput in Neural Syst 7:87-107.

Park TJ, Klug A, Holinstat M, Grothe B (2004) Interaural level difference processing in the lateral superior olive and the inferior colliculus. J Neurophysiol 92:289-301.

Parzen E (1962) On estimation of a probability density function and mode. Ann Math Statist 33:1065-1076.

Petersen RS, Panzeri S, Diamond ME (2001) Population coding of stimulus location in rat somatosensory cortex. Neuron 32:503-514.

Phillips DP, Irvine DRF (1981) Responses of single neurons in physiologically defined area AI of cat cerebral cortex: sensitivity to interaural intensity differences. Hear Res 4:299-307.

Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky EJ (2005) Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. J Neurosci 25:11003-11013.

Puchalla JL, Schneidman E, Harris RA, Berry MJ (2005) Redundancy in the population code of the retina. Neuron 46:493-504.

Ramachandran R, Davis KA, May BJ (1999) Single-unit responses in the inferior colliculus of decerebrate cats. I. Classification based on frequency response maps. J Neurophysiol 82:152-163.

Ramachandran R, May BJ (2002) Functional segregation of ITD sensitivity in the inferior colliculus of decerebrate cats. J Neurophysiol 88:2251-2261.

Reich DS, Mechler F, Victor JD (2001a) Independent and redundant information in nearby cortical neurons. Science 294:2566-2568.

Reich DS, Mechler F, Victor JD (2001b) Formal and attribute-specific information in primary visual cortex. J Neurophysiol 85:305-318.

Reich DS, Mechler F, Victor JD (2001) Temporal coding of contrast in primary visual cortex. When, what, and why. J Neurophysiol 85:1039-1050.

Reiss LJ, Young ED (2005) Spectral edge sensitivity in neural circuits of the dorsal cochlear nucleus. J Neurosci 25:3680-3691.

Rice JJ, May BJ, Spirou GA, Young ED (1992) Pinna-based spectral cues for sound localization in cat. Hear Res 58:132-152.

Richmond BJ, Optican LM (1990) Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex II. Information transmission. J Neurophysiol 64:370-380.

Rolls ET, Treves A, Tovee MJ (1997) The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. Exp Brain Res 114:149-162.

Rose JE, Brugge JF, Anderson DJ, Hind JE (1967) Phase-locked responses to low-frequency tones in single auditory nerve fibers of the squirrel monkey. J Neurophysiol 30:769-793.

Rose JE, Greenwood DD, Goldberg JM, Hind JE (1963) Some discharge characteristics of single neurons in the inferior colliculus of the cat. I. Tonotopical organization, relation of spike counts to tone intensity, and firing patterns of single elements. J Neurophysiol 26:294-320.

Rose JE, Gross NB, Geisler CD, Hind JE (1966) Some neural mechanisms in the inferior colliculus of the cat which may be relevant to localization of a sound source. J Neurophysiol 29:288-314.

Roth GL, Aitkin LM, Andersen RA, Merzenich MM (1978) Some features of the spatial organization of the central nucleus of the inferior colliculus of the cat. J Comp Neurol 182:661-680.

Roth GL, Kochar RK, Hind JE (1980) Interaural time differences: Implications regarding the neurophysiology of sound localization. J Acoust Soc Am 68:1643-1651.

Rothman JS, Young ED, Manis PB (1993) Convergence of auditory nerve fibers onto bushy cells in the ventral cochlear nucleus: Implications of a computational model. J Neurophysiol 70:2562-2583.

de Ruyter van Steveninck R, Bialek W (1988) Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. Proc R Soc Lond B 234:379-414.

Ryugo DK, Sento S (1991) Synaptic connections of the auditory nerve in cats: relationship between endbulbs of held and spherical bushy cells. J Comp Neurol 305:35-48.

Schraudolph NN (2004) Gradient-based manipulations of nonparametric entropy estimates. IEEE Trans on Neural Networks 15:828-837.

Semple MN, Kitzes LM (1987) Binaural processing of sound pressure level in the inferior colliculus. J Neurophysiol 57:1130-1147.

Shackleton TM, Skottun BC, Arnott RH, Palmer AR (2003) Interaural time difference discrimination thresholds for single neurons in the inferior colliculus of guinea pigs. J Neurosci 23:716-724.

Shannon CE (1948) A mathematical theory of communication. Bell Sys Tech J 27:379-423, 623-656.

Spirou GA, Rager J, Manis PB (2005) Convergence of auditory-nerve fiber projections onto globular bushy cells. Neuroscience 136:843-863.

Spirou GA, Young ED (1991) Organization of dorsal cochlear nucleus type IV unit response maps and their relationship to activation by bandlimited noise. J Neurophysiol 66:1750-1768.

Stark H, Woods JW (2002) Probability and random processes with applications to signal processing. Upper Saddle River, NJ: Prentice Hall.

Stecker GC, Middlebrooks JC (2003) Distributed coding of sound location in the auditory cortex. Biol Cybern 89:341-349.

Sterbing SJ, Hartung K, Hoffmann KP (2003) Spatial tuning to virtual sounds in the inferior colliculus of the guinea pig. J Neurophysiol 90:2648-2659.

Stotler WA (1953) An experimental study of the cells and connections of the superior olivary complex of the cat. J Comp Neurol 98:401-432.

Strong SP, de Ruyter van Steveninck RR, Bialek W, Koberle R (1998) On the application of information theory to neural spike trains. Pac Symp Biocomput 3:621-632.

Sutherland DP, Masterton RB, Glendenning KK (1998). Role of acoustic striae in hearing: Reflexive responses to elevated sound-sources. Behav Brain Res 97:1-12.

Teich MC (1989) Fractal character of the auditory neural spike train. IEEE Trans Biomed Eng 36:150-160.

Teich MC, Johnson DH, Kumar AR, Turcott RG (1990) Rate fluctuations and fractional power-law noise recorded from cells in the lower auditory pathway of the cat. Hear Res 46:41-52.

Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci 20:2315-2331.

Tollin DJ (2003) The lateral superior olive: a functional role in sound source localization. Neuroscientist 9:127-143.

Tollin DJ, Yin TCT (2002) The coding of spatial location by single units in the lateral superior olive of the cat. II. The determinants of spatial receptive fields in azimuth. J Neurosci 22:1468-1479.

Treves A, Panzeri S (1995) The upward bias in measures of information derived from limited data samples. Neural Comput 7:399-407.

Victor JD (2002) Binless strategies for estimation of information from neural data. Phys. Rev. E 66:051903.

Victor JD, Purpura KP (1997) Metric-space analysis of spike trains: theory, algorithms and application. Network: Comput in Neural Syst 8:127-164.

Webster DR, Popper AN, Fay RR (1992) The mammalian auditory pathway: neuroanatomy. Springer-Verlag: New York.

Weiss TF (1964) A model for firing patterns of auditory nerve fibers. Tech Rept MIT Res Lab Electron 418.

Wenstrup JJ, Fuzessery ZM, Pollak GD (1988) Binaural neurons in the mustache bat's inferior colliculus. I. Responses of 60-kHz EI units to dichotic sound stimulation. J Neurophysiol 60:1369-1383.

Wiener MC, Richmond BJ (2003) Decoding spike trains instant by instant using order statistics and the mixture-of-Poissons model. J Neurosci 23:2394-2406.

Wightman FL, Kistler DJ (1992) The dominant role of low-frequency interaural time differences in sound localization. J Acoust Soc Am 91:1648-1661.

Winter IM, Palmer AR (1995) Level dependence of cochlear nucleus onset unit responses and facilitation by second tones or broadband noise. J Neurophysiol 73:141-159.

Wise LZ, Irvine DRF (1985) Topographic organization of interaural intensity difference sensitivity in deep layers of cat superior colliculus: Implications for auditory spatial representation. J Neurophysiol 54:185-211.

Wu SH, Oertel D (1984) Intracellular injection with horseradish peroxidase of physiologically characterized stellate and bushy cells in slices of mouse anteroventral cochlear nucleus. J Neurosci 4:1577-1588.

Yin TCT (1994) Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat. J Neurosci 14:5170-5186.

Yin TCT, Chan JCK (1990) Interaural time sensitivity in medial superior olive of cat. J Neurophysiol 64:465-88.

Young ED, Brownell WE (1976) Responses to tones and noise of single cells in the dorsal cochlear nucleus of unanesthetized cats. J Neurophysiol 39:282-300.

Young ED, Davis KA (2002) Circuitry and function of the dorsal cochlear nucleus. In: Integrative functions in the mammalian auditory pathway (Oertel D, Fay RR, Popper AN, eds.), pp160-206. New York: Springer.

Young ED, Spirou GA, Rice JJ, Voigt HF (1992) Neural organization and responses to complex stimuli in the dorsal cochlear nucleus. Philos Trans R Soc Lond Biol Sci 336:407-413.

Yu JJ, Young ED (20000 Linear and nonlinear pathways of spectral information transmission in the cochlear nucleus. Proc Natl Acad Sci USA 97:11780-11786.

Zhang X, Heinz MG, Bruce IC, Carney LH (2001) A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. J Acoust Soc Am 109:648-670.

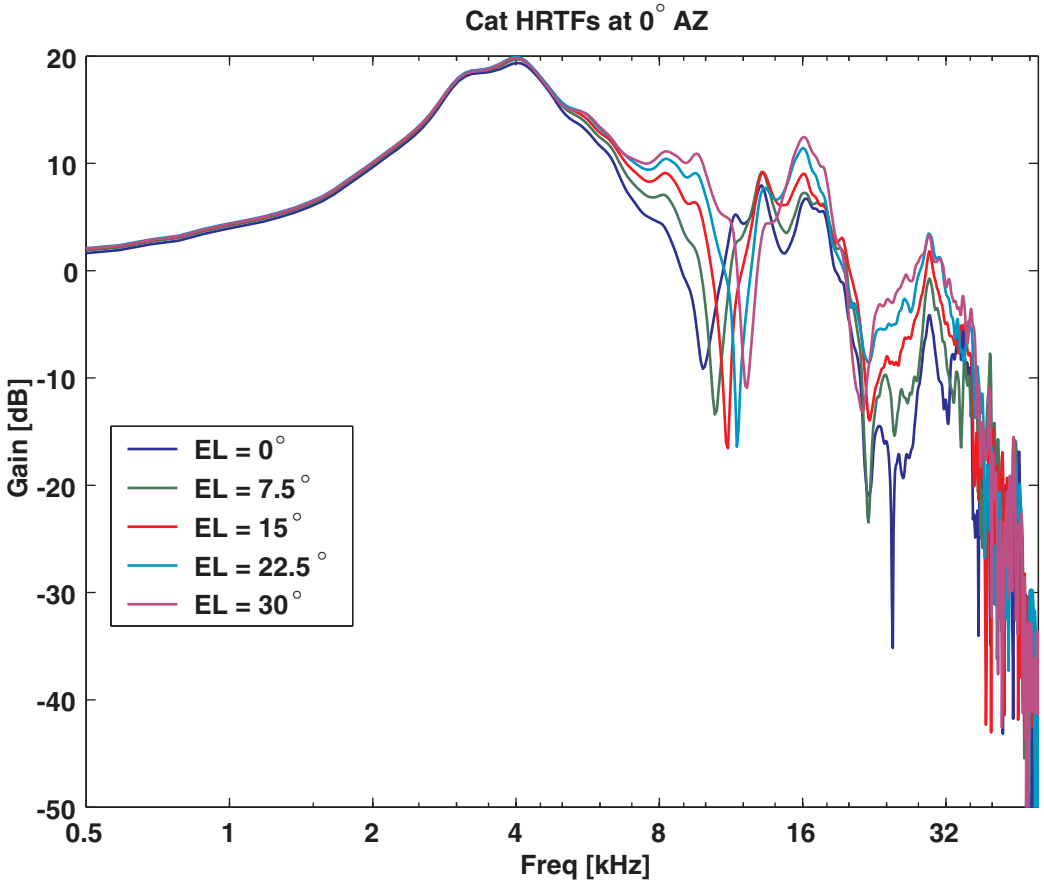


Figure 1.1: Example HRTFs measured in cats. Gain is plotted as a function of frequency for five speaker locations in the mid-sagittal plane. Note the prominent spectral notch between 9 and 12 kHz.

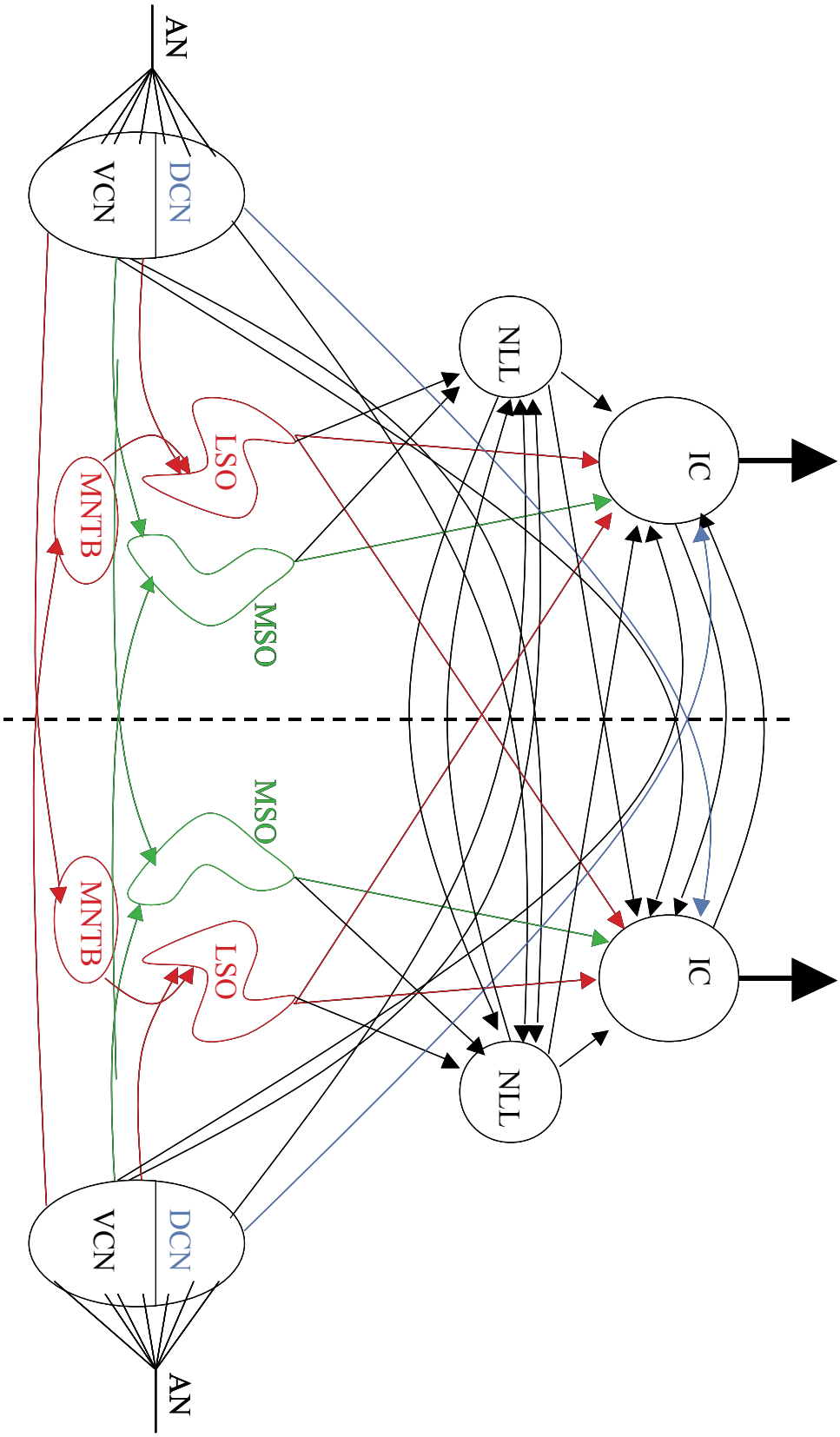

AN
VCN
DCN
NLL
IC
LSO
MNTB
MSO
MSO
MNTB
LSO
IC
NLL
VCN
DCN
AN

Figure 1.2: Schematic of the auditory brainstem circuitry up to the level of the inferior colliculus. *Green:* primary ITD processing pathway. *Red:* primary ILD processing pathway. *Blue:* primary SN processing pathway.

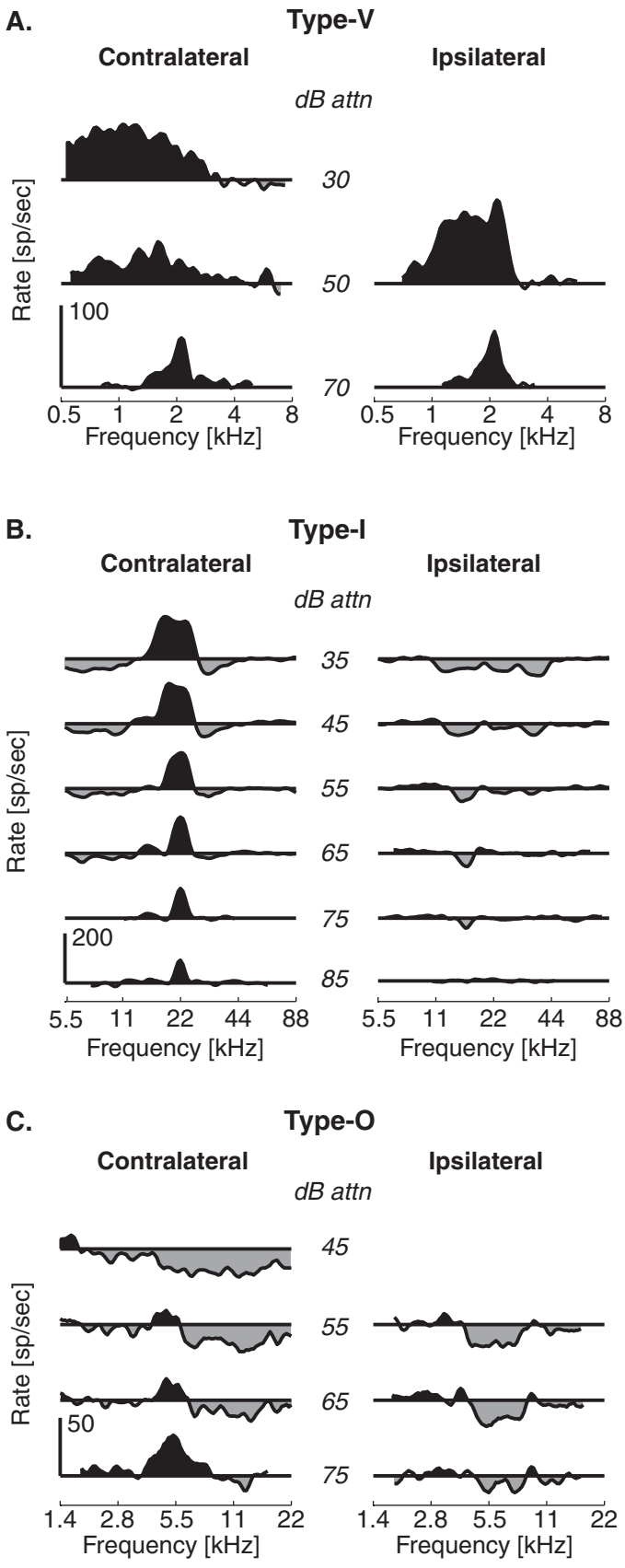

A.
Type-V
Contralateral
Ipsilateral
dB attn
30
50
70
100
Rate [sp/sec]
0.5 1 2 4 8
Frequency [kHz]
B.
Type-I
Contralateral
Ipsilateral
dB attn
35
45
55
65
75
85
200
Rate [sp/sec]
5.5 11 22 44 88
Frequency [kHz]
C.
Type-O
Contralateral
Ipsilateral
dB attn
45
55
65
75
50
Rate [sp/sec]
1.4 2.8 5.5 11 22
Frequency [kHz]

Figure 1.3: Response maps typical of 3 of the 4 physiological classes of IC neurons, those exhibiting sustained responses to tones. Response maps are calculated as the driven rate to a 200 ms pure tone presented to one ear, with the spontaneous rate subtracted to reveal inhibition (in grey). At 0 dB attenuation, the sound level is approximately 100 dB SPL. **A.** Type-V neuron. **B.** Type-I neuron. **C.** Type-O neuron.

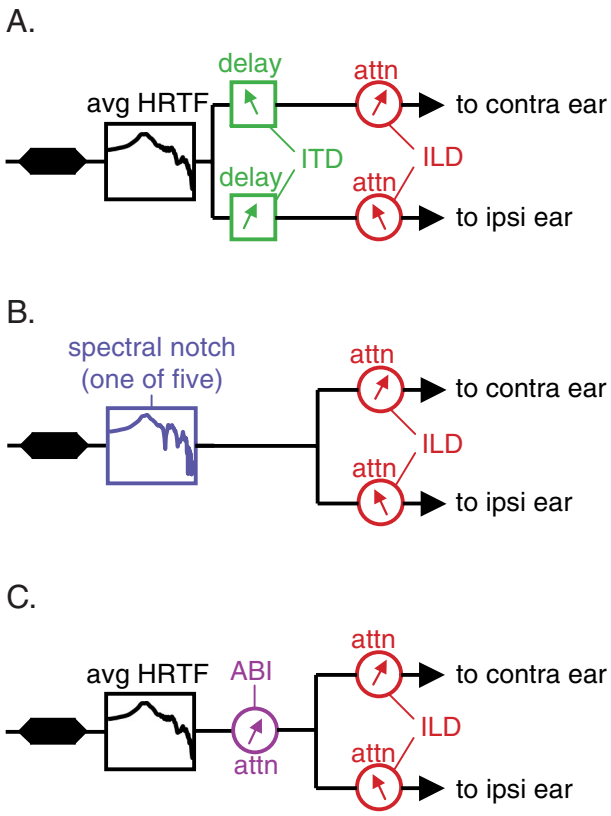


Figure 2.1: Stimulus construction. **A.** ITD/ILD stimulus set: Frozen noise was filtered through a cat HRTF, averaged over spatial location to give the location-independent spectral characteristics of the pinnae and ear canal. The noise was then copied into two streams, on which an ITD and an ILD were placed. **B.** SN/ILD stimulus set: The frozen noise was filtered through one of five HRTFs showing a prominent spectral notch, split into two streams, and attenuated relative to one another. The five HRTFs used are shown in Fig. 1.1. **C.** ABI/ILD stimulus set: As in A., the frozen noise was filtered through the spatially averaged HRTF. It was then attenuated to one of five levels before being split into two streams that were attenuated relative to one another. The ILD was imposed such that the average binaural intensity was preserved.

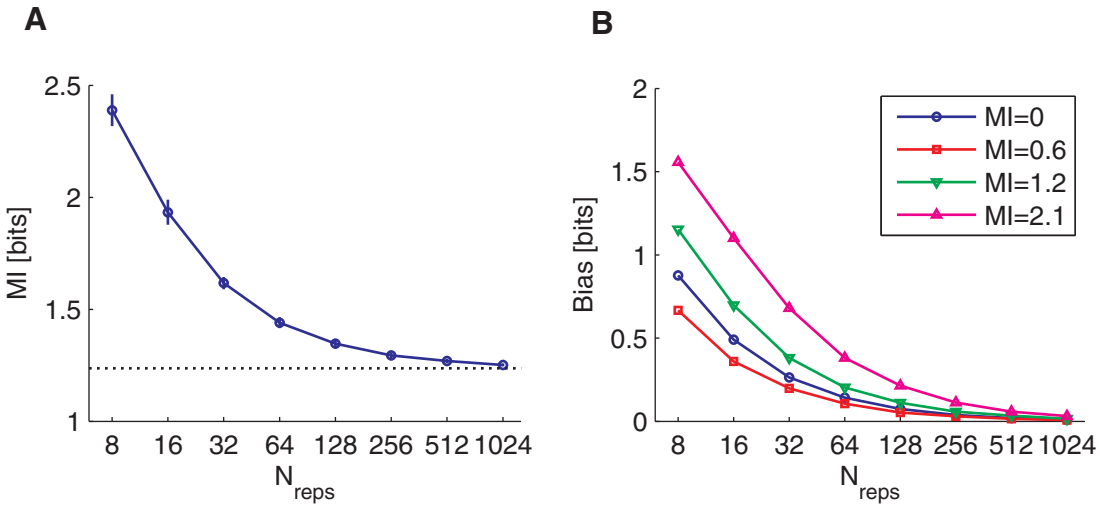


Figure 3.1: Bias in MI estimates made with the MLE. **A.** The convergence of the maximum likelihood estimates of MI as a function of the amount of data for one simulation, with $\lambda(s)=s$, where s is the stimulus index and runs 1 to 25. The real MI in the system is ~1.2 bits, indicated by the dotted black line. The error bars show the mean (± 1 SD) of 50 runs of the simulation for every sample size. For $N_{reps}>64$, the error bars are smaller than the symbol size. **B.** The bias of the MLE as a function of N_{reps}, for several simulations with different information contents. The stimulus parameters were: for MI=0, $\lambda(s)=5$ for all stimuli; for MI=0.6, $\lambda(s)=s/4$; for MI=1.2, $\lambda(s)=s$; and for MI=2.1, $\lambda(s)=4s$.

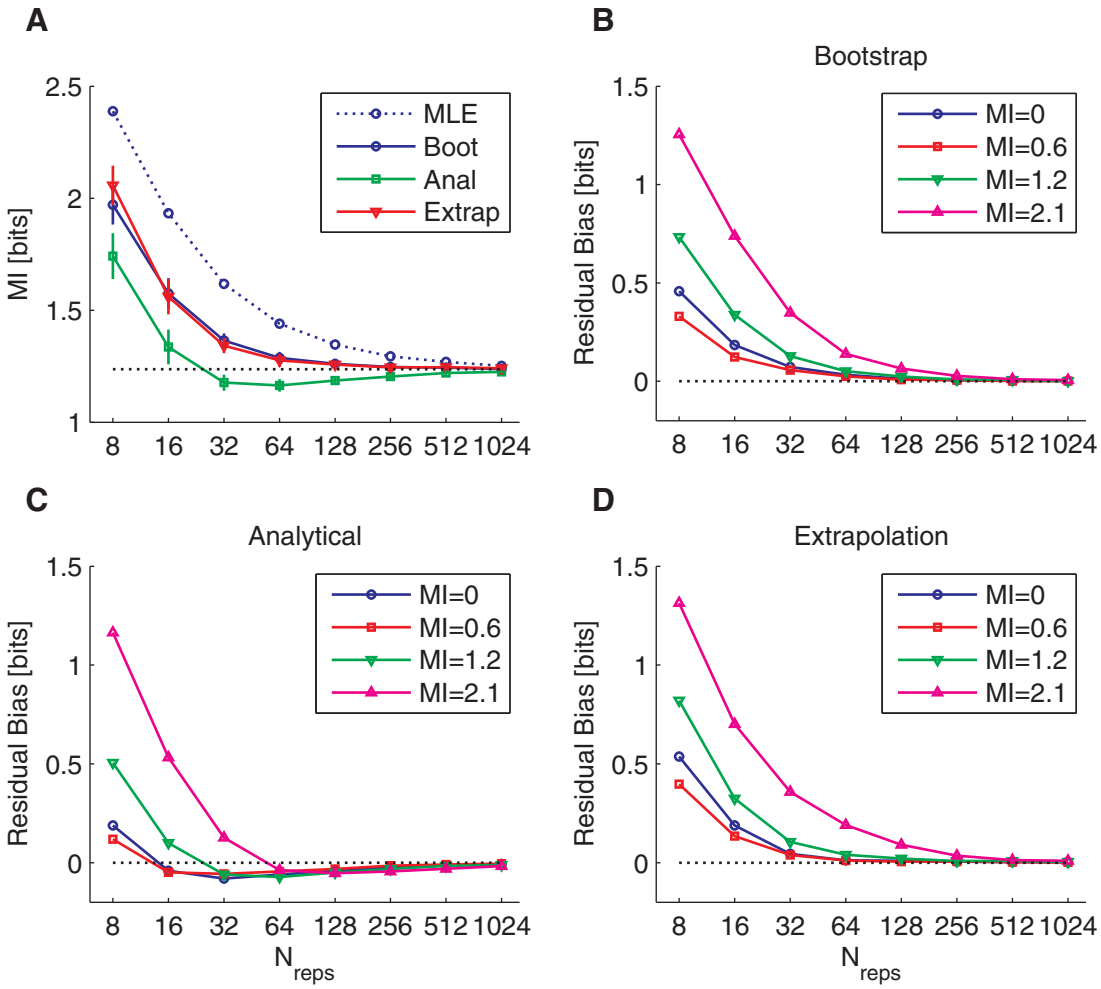


Figure 3.2: A comparison of debiasing techniques for the spike count MI. **A.** Debiased estimates of MI are plotted as a function of sample size for several debiasing techniques. The dotted black line represents the real MI. **B.** The residual bias is plotted as a function of N_{reps} for MI estimates debiased with the bootstrap method, for several different values of the real MI. **C, D.** Same as B, for the analytical and extrapolotation methods. All simulation parameter values are the same as in Fig. 3.1.

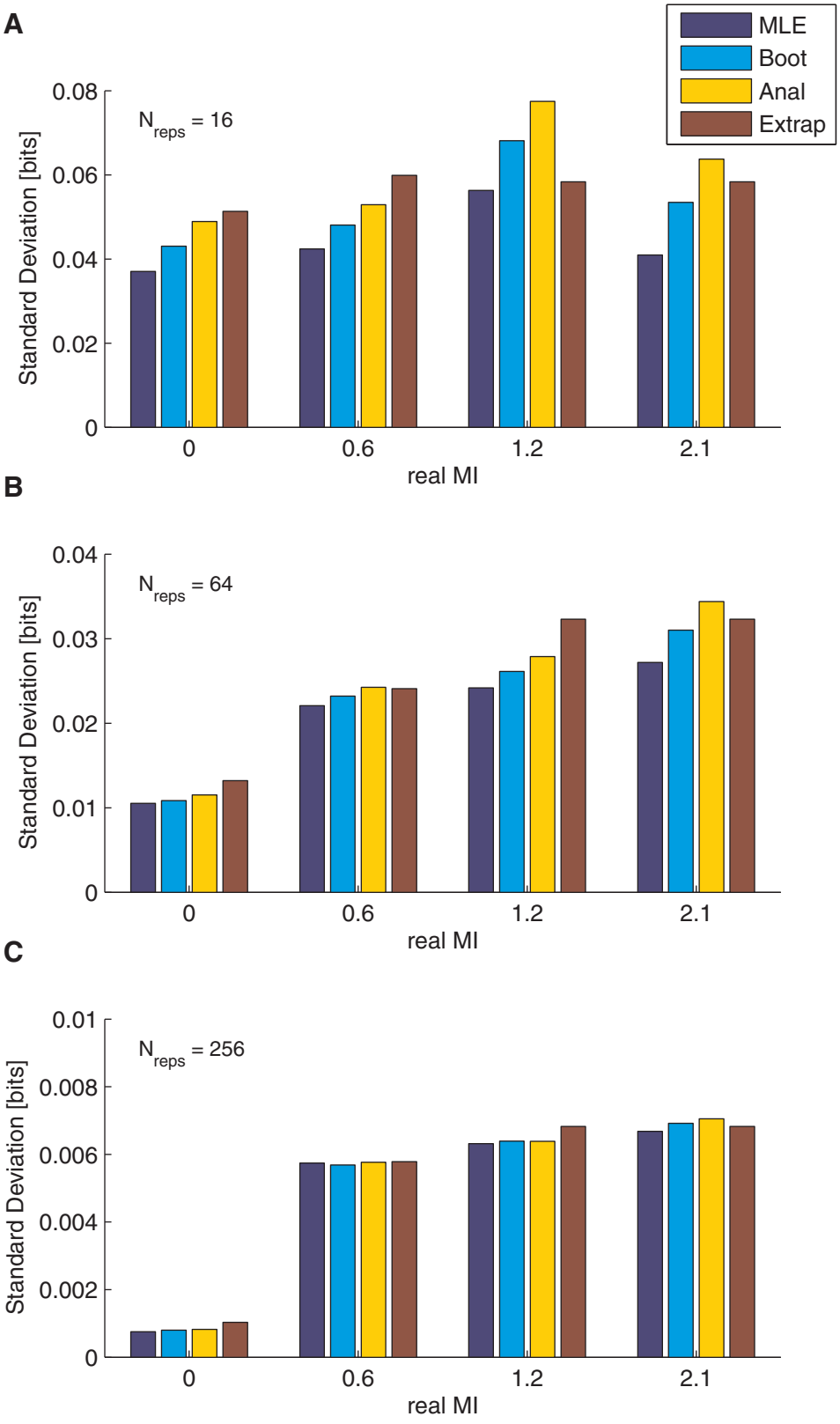

A
MLE
Boot
Anal
Extrap
0.08
0.06
0.04
0.02
0
N reps = 16
Standard Deviation [bits]
0
0.6
1.2
2.1
real MI
B
0.04
0.03
0.02
0.01
0
N reps = 64
Standard Deviation [bits]
0
0.6
1.2
2.1
real MI
C
0.01
0.008
0.006
0.004
0.002
0
N reps = 256
Standard Deviation [bits]
0
0.6
1.2
2.1
real MI

Figure 3.3: A comparison of the standard deviations in debiased MI estimates made with the MLE. **A.** Bar plot of the standard deviations in MI estimates for various estimation techniques, grouped by the real MI value. For all measurements, N_{reps}=16 **B, C.** Same as A., for N_{reps} = 64 and N_{reps}=256, respectively.

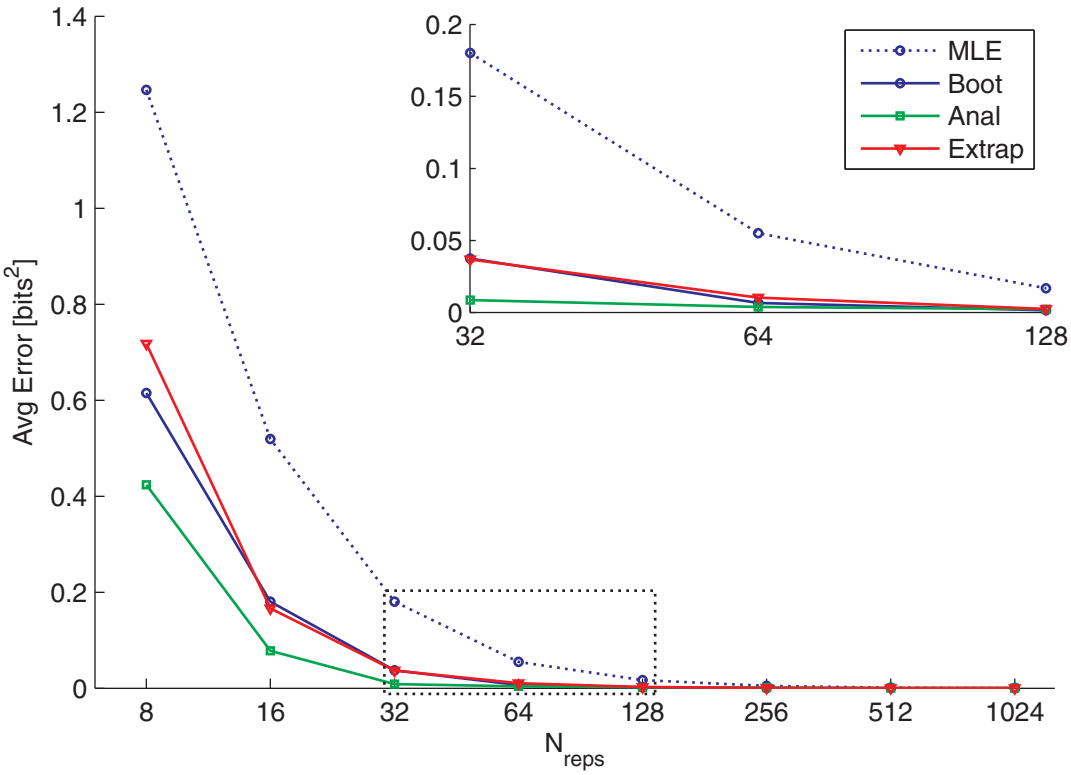


Figure 3.4: Average error in MI estimates. Error is plotted as a function of N_{reps}, averaged over the four simulations run with different real MI values (described in Fig. 3.1). The inset shows a close up of the area indicated with the dotted black box.

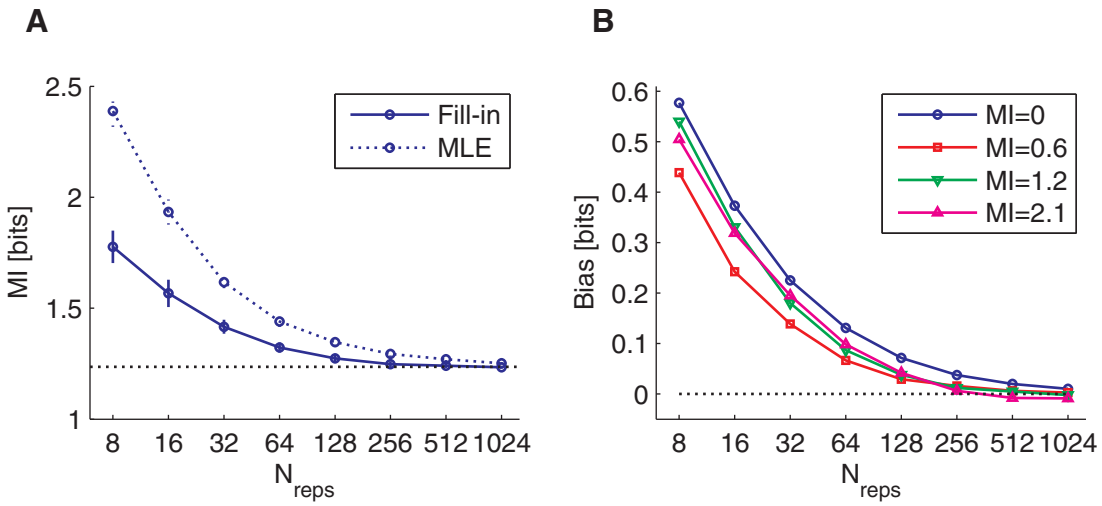


Figure 3.5: Bias in fill-in estimates of spike count MI. **A.** Convergence of fill-in MI estimates with sample size, compared to the convergence of MI estimates with the MLE method. **B.** The bias of the fill-in MI estimates is plotted as a function of N_{reps} for several different information contents. Simulation parameters are the same as for Fig. 3.1.

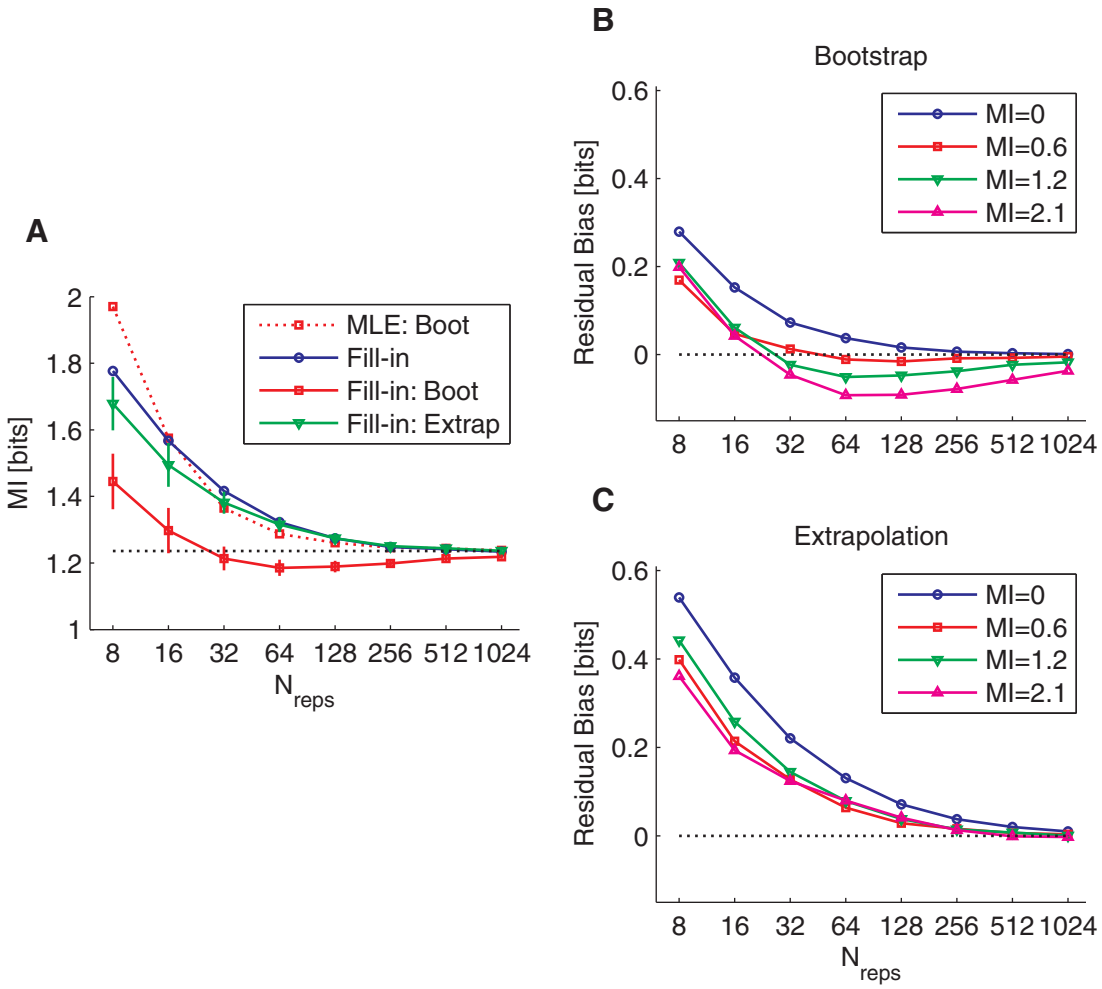


Figure 3.6: Comparison of debiasing techniques applied to fill-in estimates of MI. **A.** Debiased fill-in estimates of MI are plotted as a function of N_{reps}. The dotted red line indicates bootstrapp debiased estimates of MI made with the MLE method, for reference. **B.** The residual bias left after debiasing the fill-in estimates with bootstrap. **C.** Same as B., for the extrapolation method.

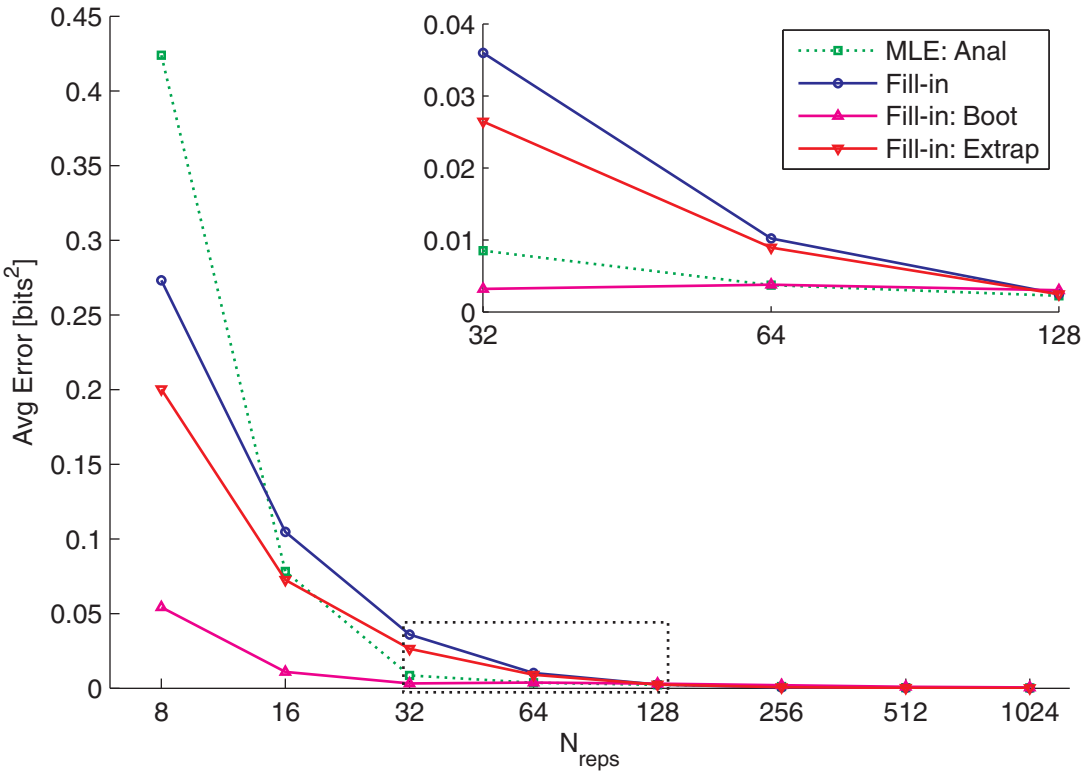


Figure 3.7: Average error in fill-in estimates of MI. Error is plotted as a function of Nreps, averaged over the four simulations run with different real MI values, as in Fig. 3.4. Error in MI estimated with the MLE method and debiased with the analytical method are shown as a dotted green line, for reference. The inset shows a close up of the area indecated with the dotted black box.

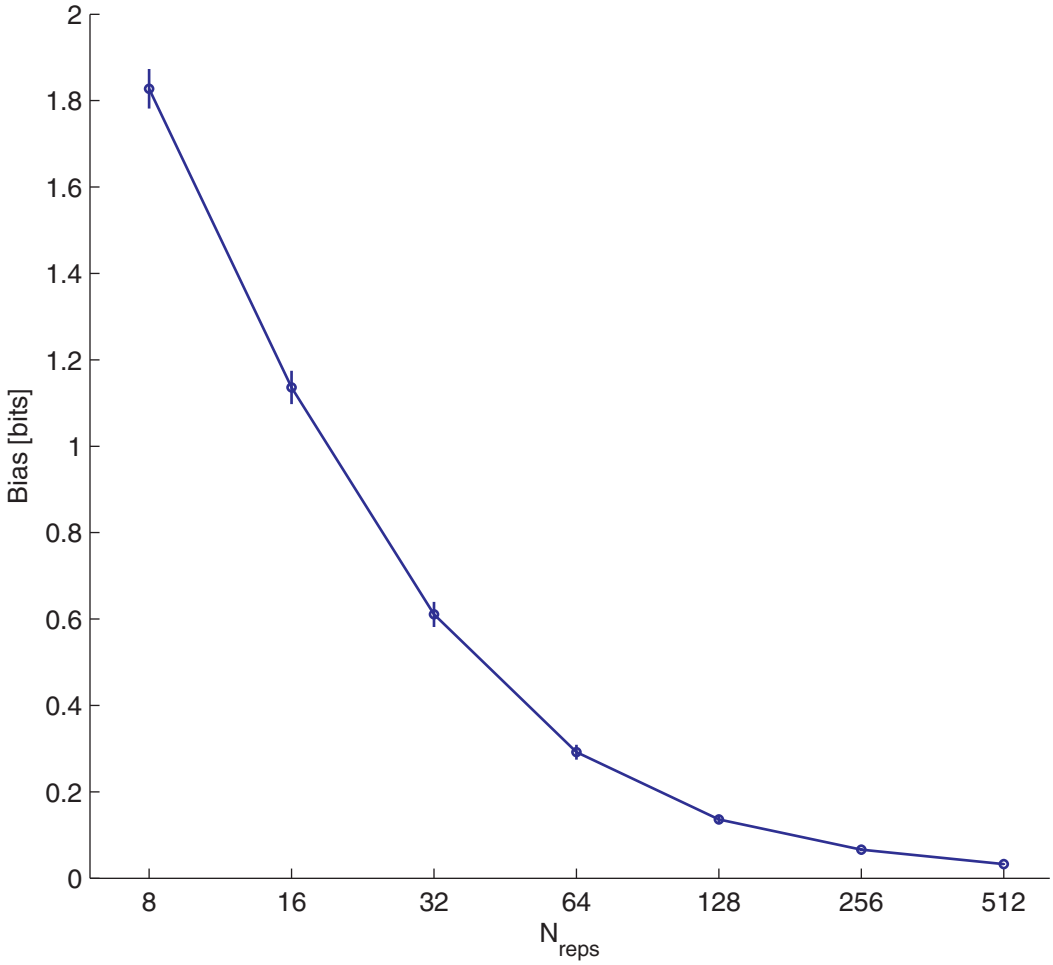


Figure 3.8: Random renormalization estimates of bias. Bias estimates are based on a 25x25 confusion matrix with N_{reps} entries per row. Points represent the mean (±1 SD) MI of 500 randomly generated confusion matrices.

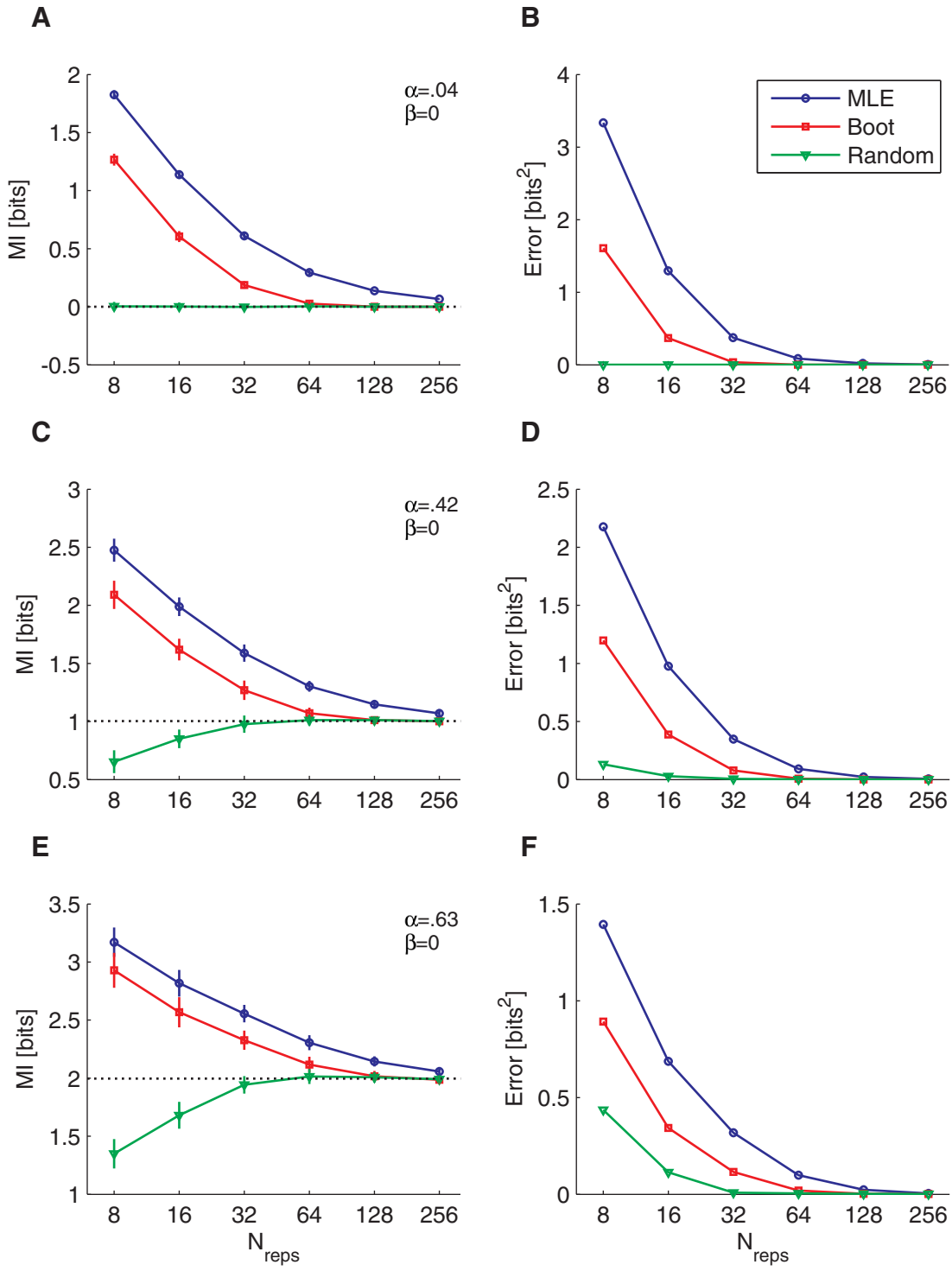

A
B
C
D
E
F
α=.04
β=0
α=.42
β=0
α=.63
β=0
MLE
Boot
Random
MI [bits]
Error [bits^2]
N$_{reps}$
8
16
32
64
128
256

Figure 3.9: A comparison of debiasing techniques for MI estimates on confusion matrices with $\beta = 0$. **A.** Convergence of MI estimates as a function of N_{reps}. The real MI value is denoted by the dotted black line. Each point represents the mean and SD of estimates from 50 independently drawn confusion matrices with the parameters specified in the upper left corner. **B.** Estimation error as a function of N_{reps} for the plots in A. **C, D.** Same as A and B, for MI = 1 bit. **E, F.** Same as A and B, for MI = 2 bits.

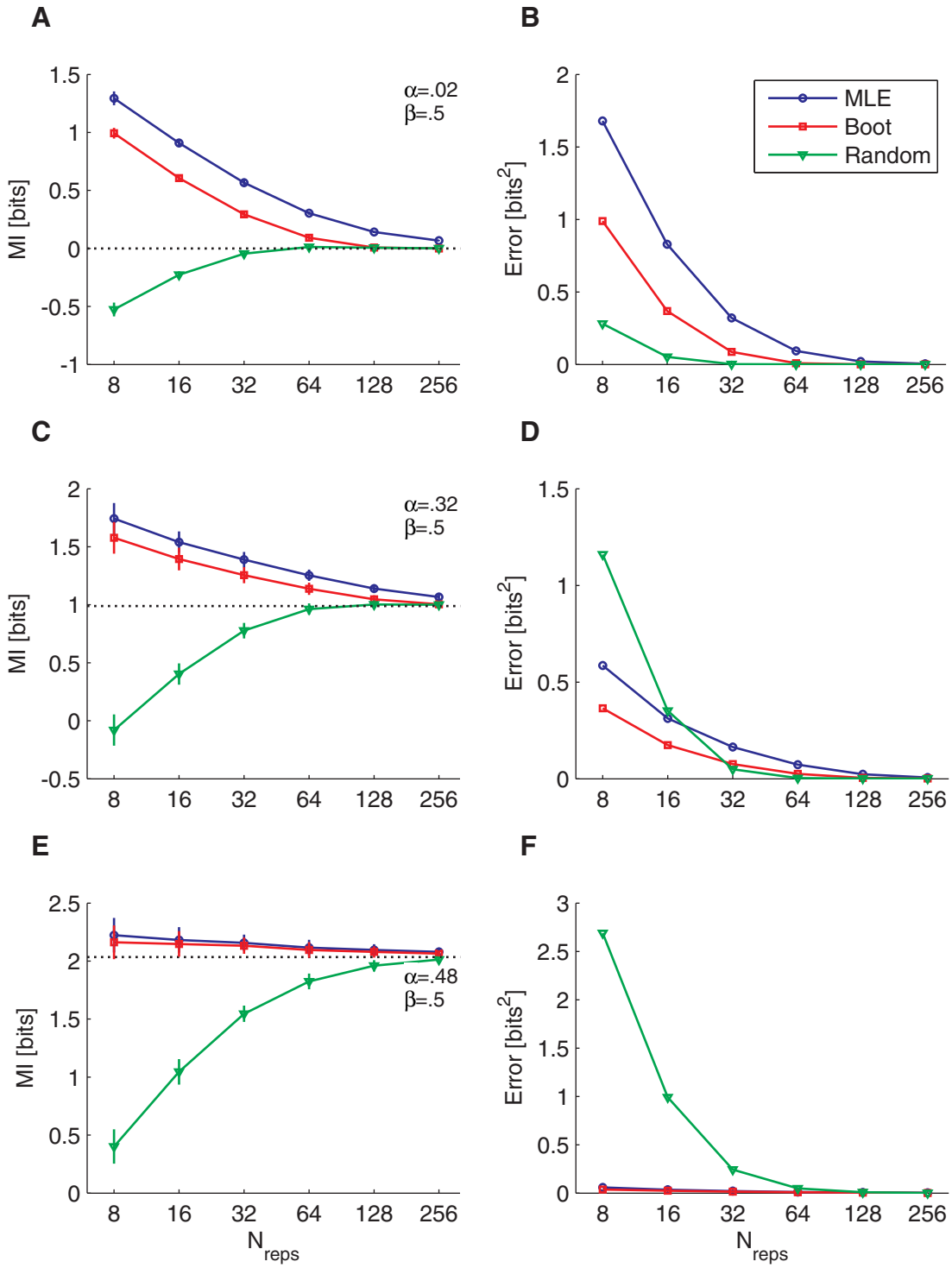


Figure 3.10: A comparison of debiasing techniques for MI estimates on confusion matrices with $\beta = 0.5$ Format is the same as Fig. 3.9.

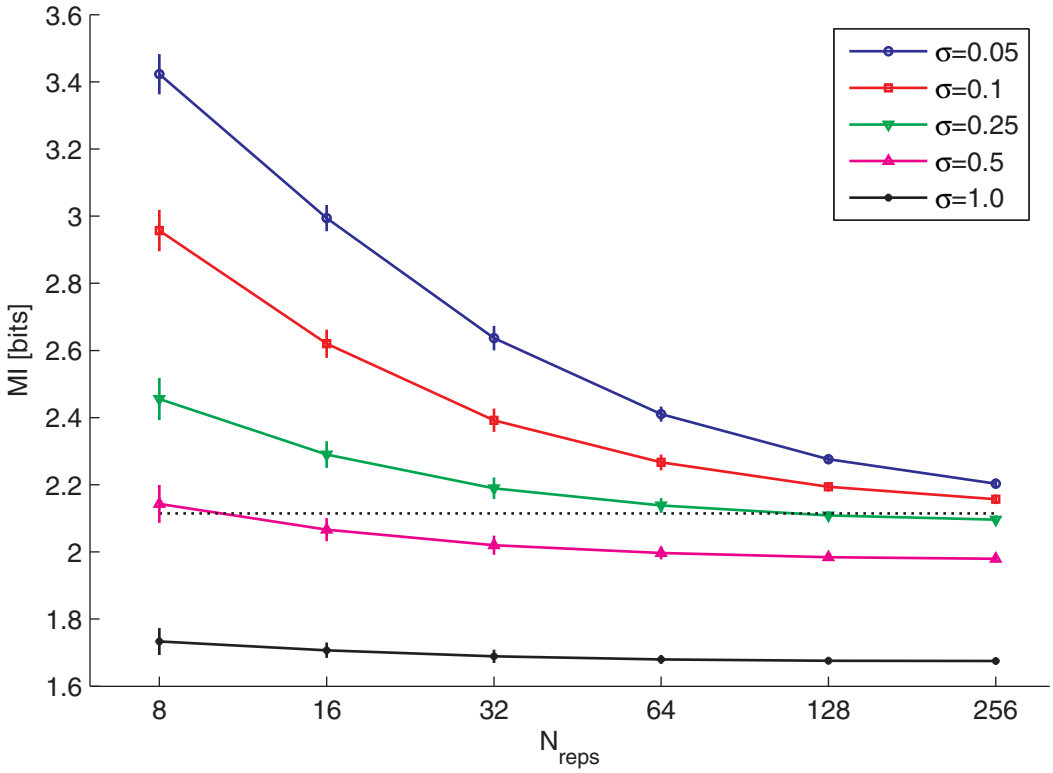


Figure 3.11: Convergence of spike timing estimates of MI made with the KDE method with several fixed kernel widths. Each point represents the mean and SD of 50 independent simulation runs.

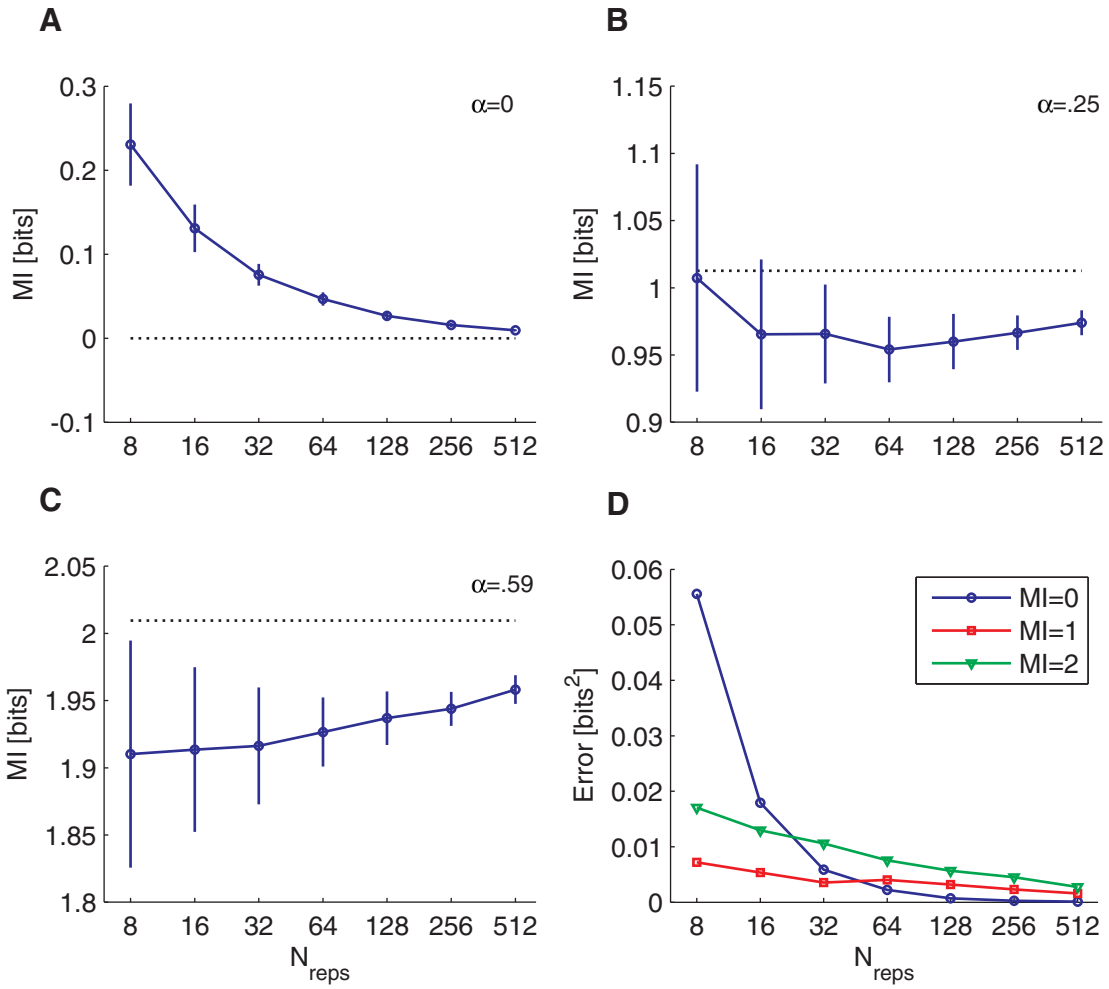


Figure 3.12: Convergence of KDE spike timing MI estimates with the optimal kernel width. **A.** Mean and standard deviation of 50 MI estimates are plotted as a function of sample size. The dotted black line represents the real MI. **B.** Same as A, with the scale factor (shown at top left) changed to give a real MI of ~1 bit. **C.** Same as A and B, for a real MI of ~2 bits. **D.** Error in the MI estimates is plotted as a function of N_{reps} for the curves in A, B, and C.

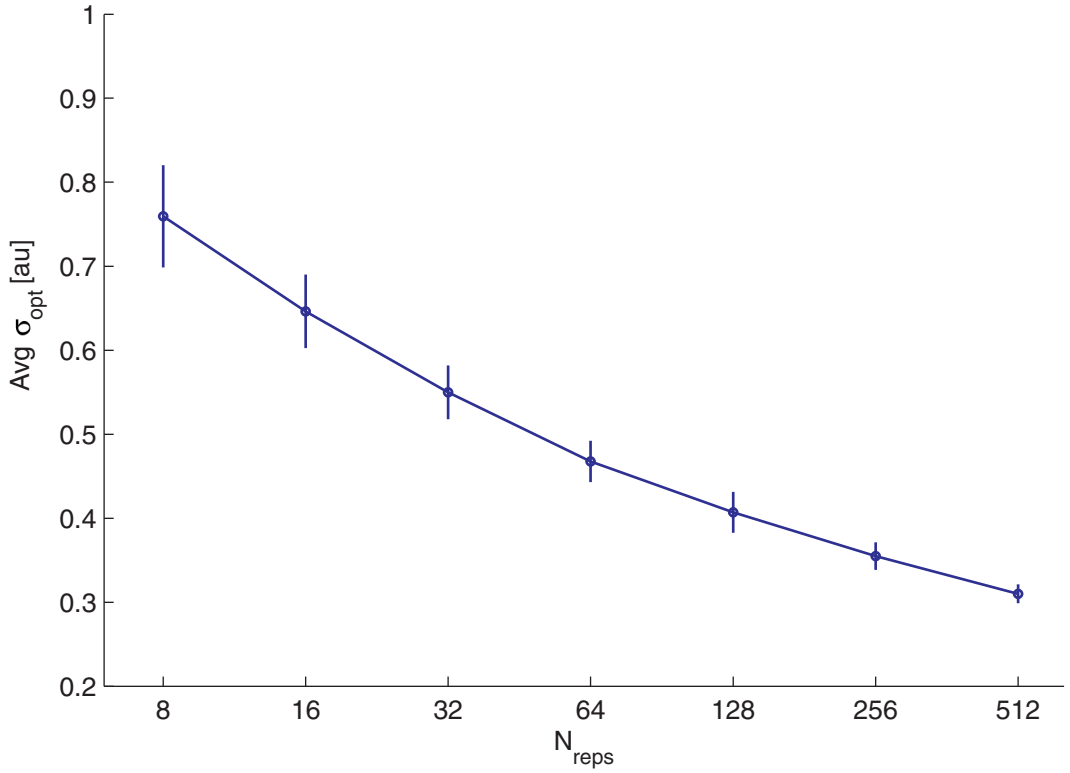


Figure 3.13: Average optimal kernel width as a function of sample size. Every kernel density estimate returns one computed optimal kernel width for each stimulus. These are averaged to give an average optimal kernel width for the simulation. Error bars show the mean and standard deviations of the average optimal kernel widths for the 50 simulations.

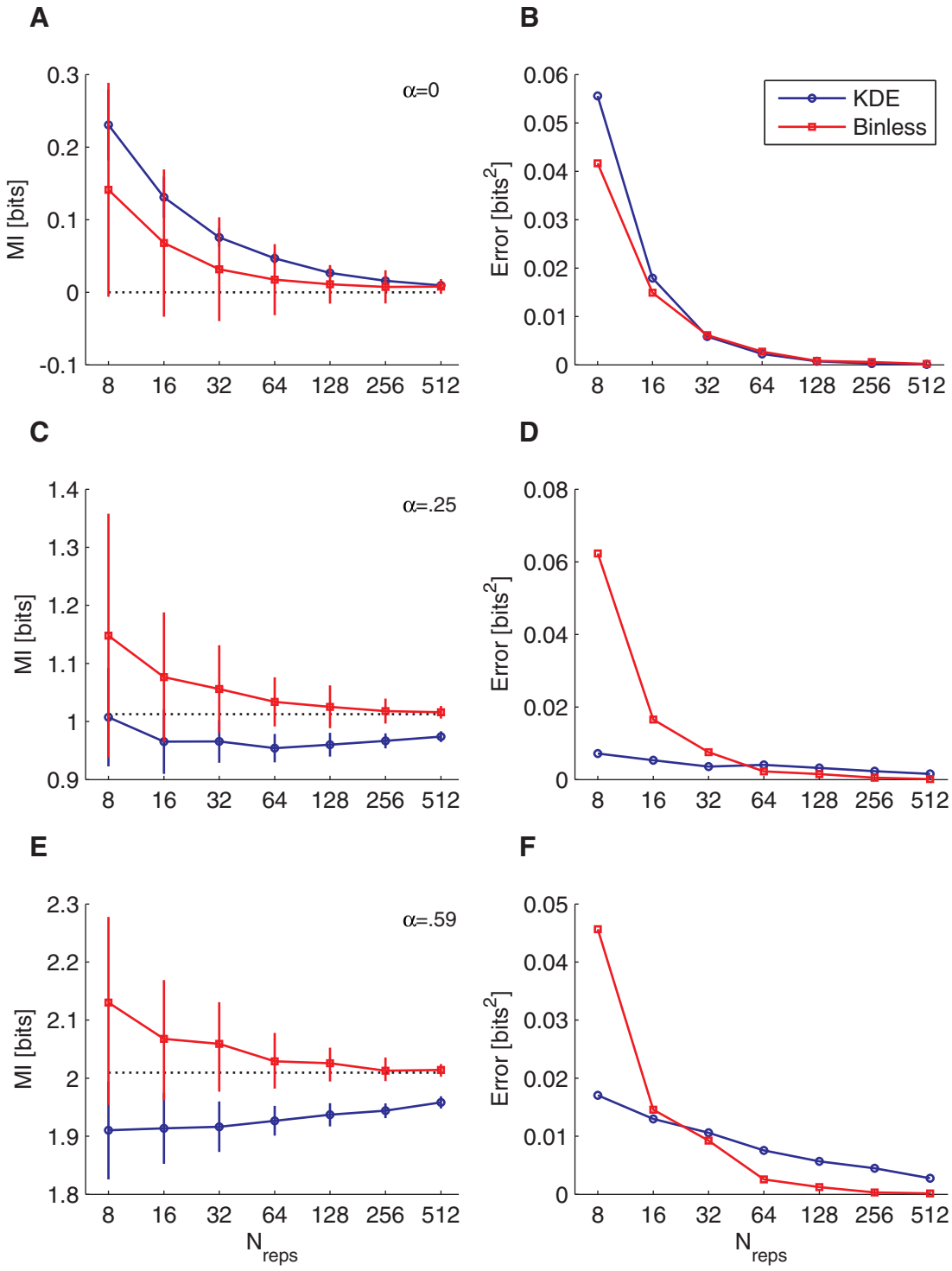

A
B
C
D
E
F
α=0
α=.25
α=.59
KDE
Binless
MI [bits]
Error [bits^2]
N$_{reps}$
8
16
32
64
128
256
512

Figure 3.14: A comparison of the binless and KDE methods of estimating MI in spike timing. **A.** Convergence of MI estimates as a function of N_{reps}. The real MI value is denoted by the dotted black line. Error bars represents the mean and SD of estimates from 50 simulations. The KDE curve is the same as that in Fig. 3.12A. **B.** Estimation error as a function of N_{reps} for the plots in A. **C, D.** Same as A and B, for MI = 1 bit. **E, F.** Same as A and B, for MI = 2 bits.

	Neural type					
Stimulus set	**V**	**I**	**O**	**onset (N)**	**other (X)**	**Total**
ITD/ILD	14	7	7	3	7	38
SN/ILD	13	40	27	7	11	98
ABI/ILD	11	15	13	3	2	44
Random Noise	2	4	7	0	1	14
Total	**29**	**50**	**40**	**10**	**20**	

Table 4.1: Counts of the numbers of neurons recorded with each stimulus set, broken down as a function of cell type. The bottom row gives the total number of neurons of each type; these counts do not equal the sums of the overlying columns because sometimes more than one stimulus set was presented to a single neuron.

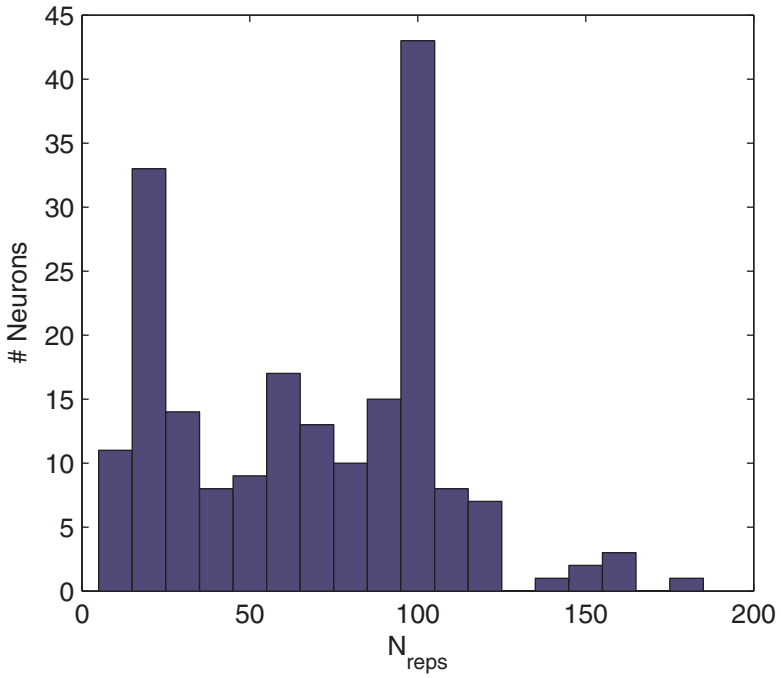


Figure 4.1: Histogram of stimulus repetitions. Each bar represents the number of neurons for which N_{reps} stimulus presentations were recorded for one stimulus set. Bin width is 10 repetitions.

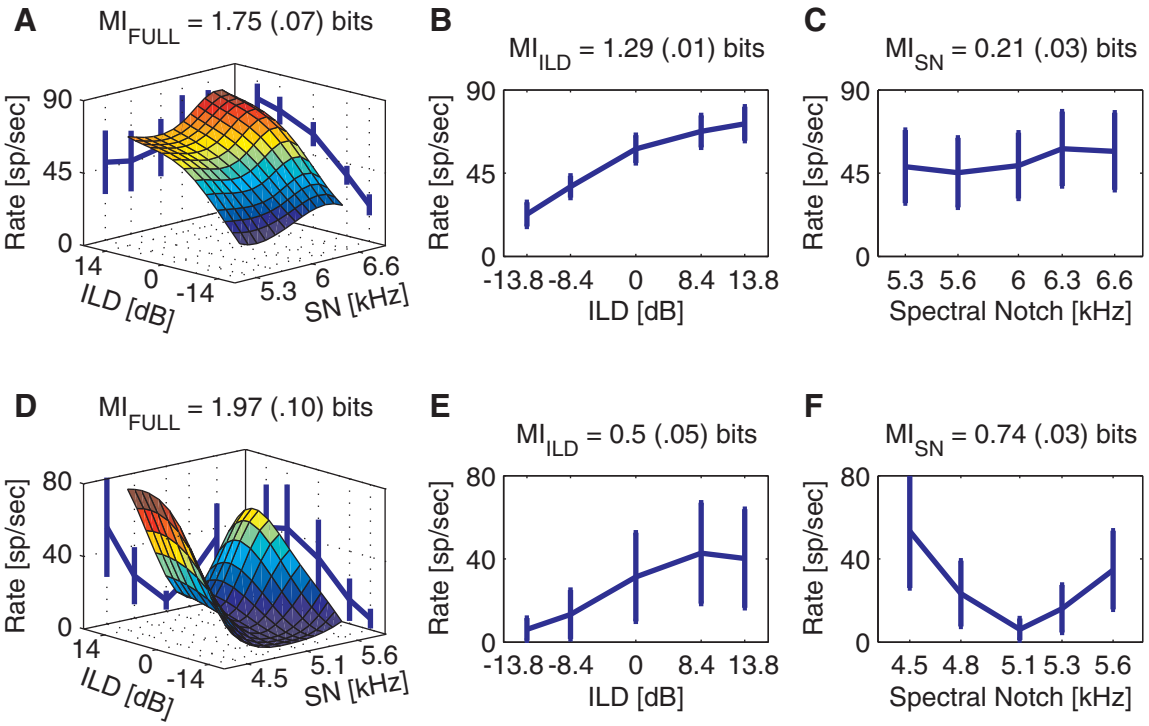


Figure 4.2: Extracting the MI to the full stimulus set and to the individual localization cues. **A.** The surface shows the mean response rate of a 6 kHz type-I neuron as a function of ILD and SN. (Surface smoothed with cubic spline interpolation). Contours on the vertical walls of the plot show the mean (±1 SD) of the rate as a function of ILD or SN only. The debiased MI is given above the plot with the bias estimate in parentheses. **B.** Mean rate v. ILD. **C.** Mean rate v. SN. **D,E,F.** Same, for another type-I neuron (BF, 5.1 kHz) showing more SN sensitivity.

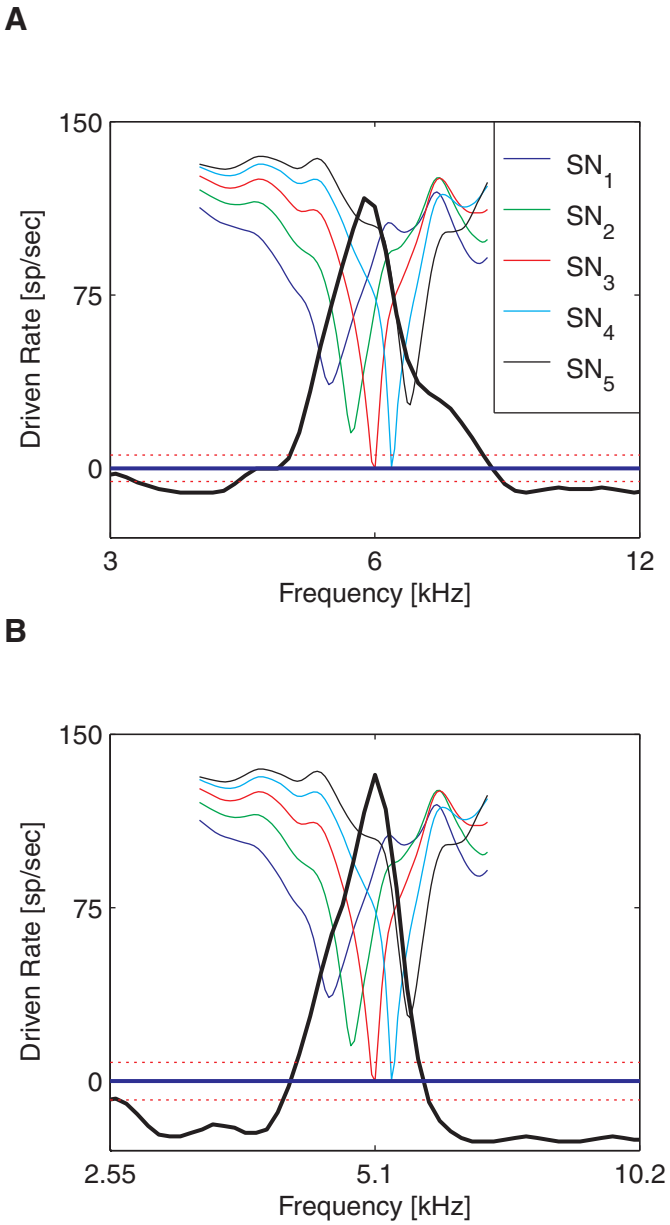


Figure 4.3: SNs within type-I response maps. The response maps (thick black lines) depict the driven rate in response to pure tones as a function of tone frequency, for the two neurons of Fig. 4.2. The dotted red lines are set at 2 SDs of the spontaneous rate. The thin lines depict the SN profiles of the five HRTFs used to construct the SN/ILD stimuli; the profiles have been truncated to show the spectra in the relevant frequency range only. The y-axis scaling of the SNs is arbitrary, these plots are meant to compare the width and spacing of the SNs to the width of the receptive field. **A.** Response map for the neuron of Fig. 4.2A-C, at a level comparable to the levels of the SNs (~40 dB above threshold). **B.** Same as A, for the neuron of Fig. 4.2D-F.

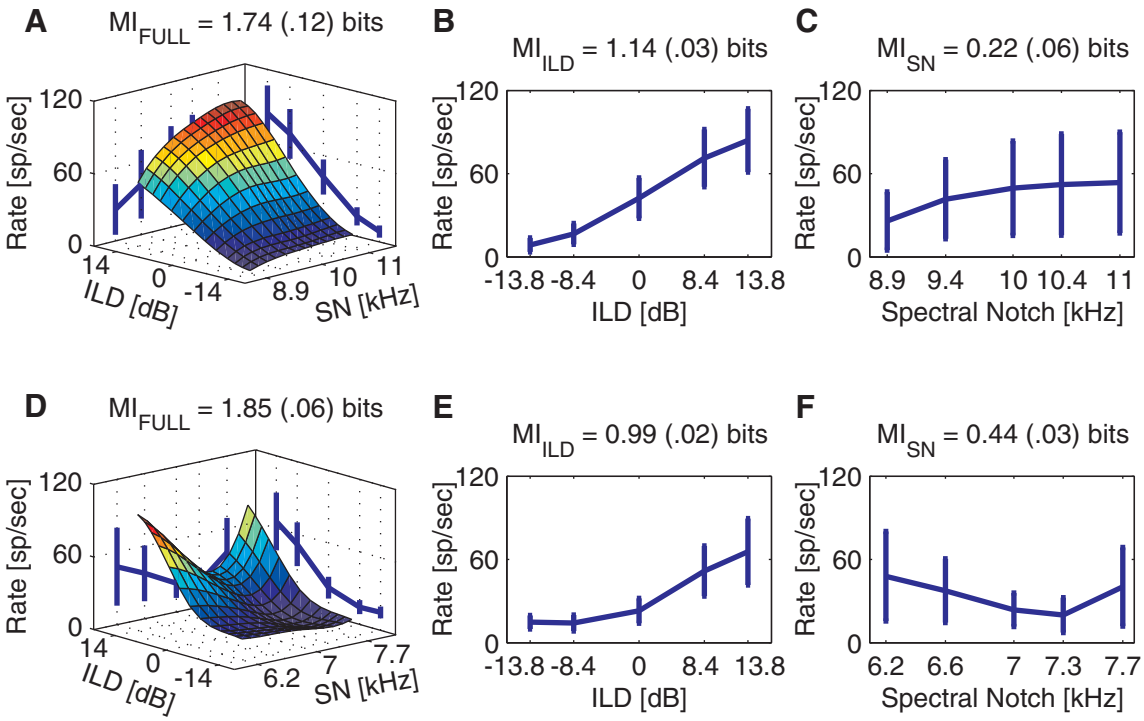


Figure 4.4: Examples of type O rate responses to the SN/ILD stimulus set. The format is the same as for Fig. 4.2 **A.** Rate surface for a 10 kHz type-O neuron. **B.** Mean rate v. ILD. **C.** Mean rate v. SN. **D,E,F.** Same, for a 7 kHz type-O neuron.

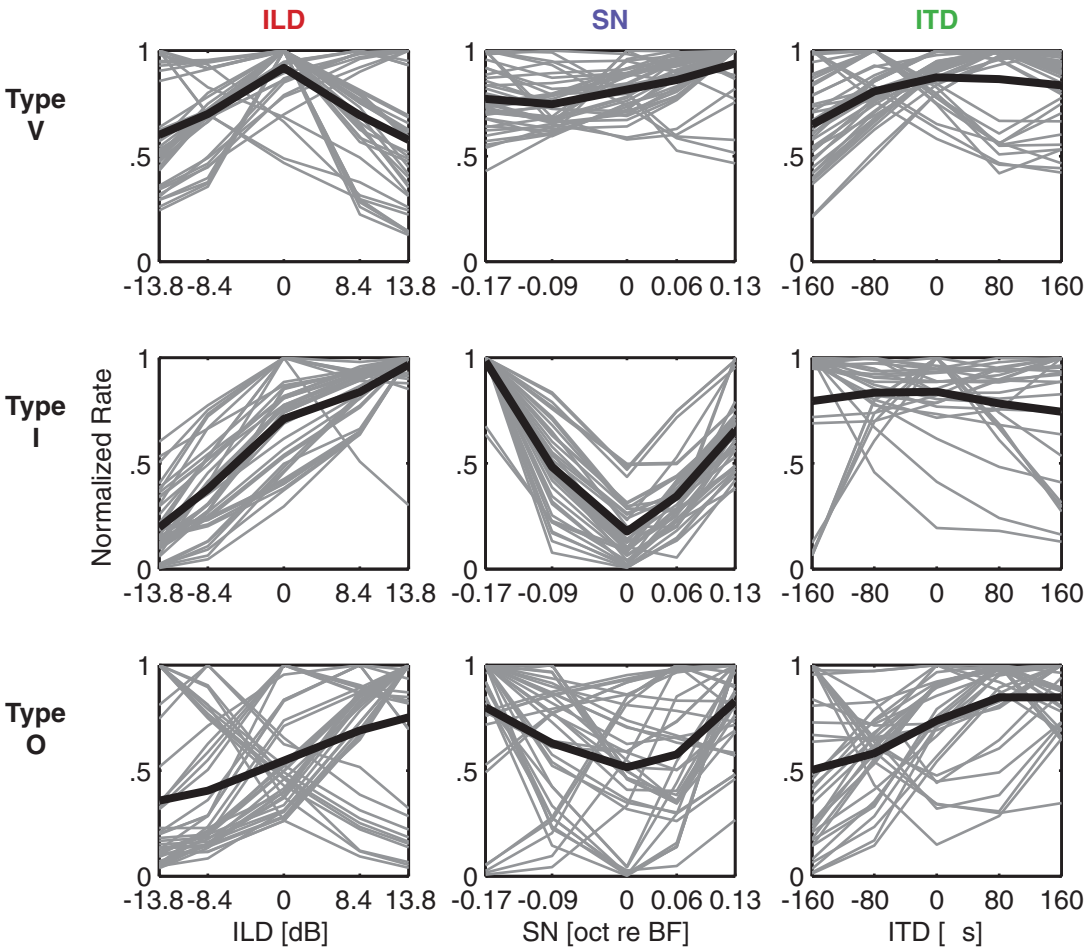


Figure 4.5: Rate profiles of responses to the SN/ILD and SN/ITD stimulus sets. Each profile has been normalized to its maximum response. Thin gray lines represent individual rate profiles, thick black lines represent the means of all rate profiles in each category.

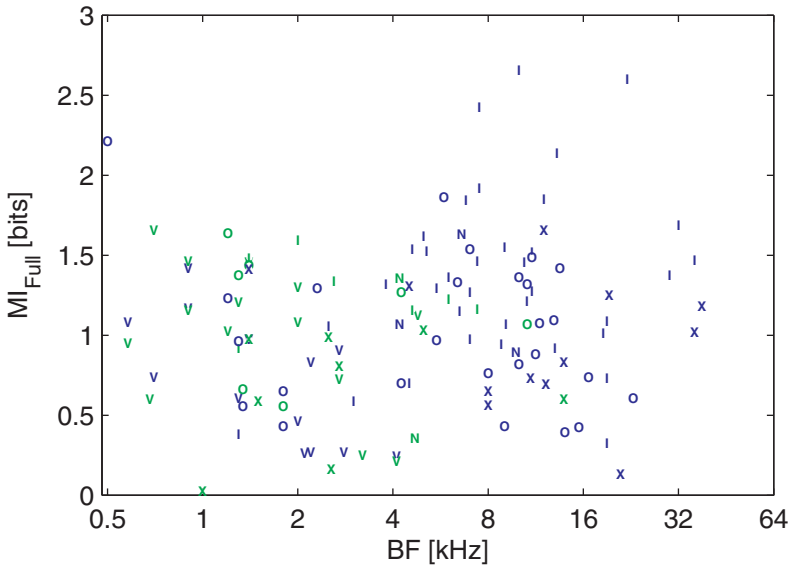


Figure 4.6: Full rate MI as a function of BF. Symbols denote the physiological cell type, and correspond to the symbols in Table 4.1. *Green:* ITD/ILD stimuli. *Blue:* SN/ILD stimuli.

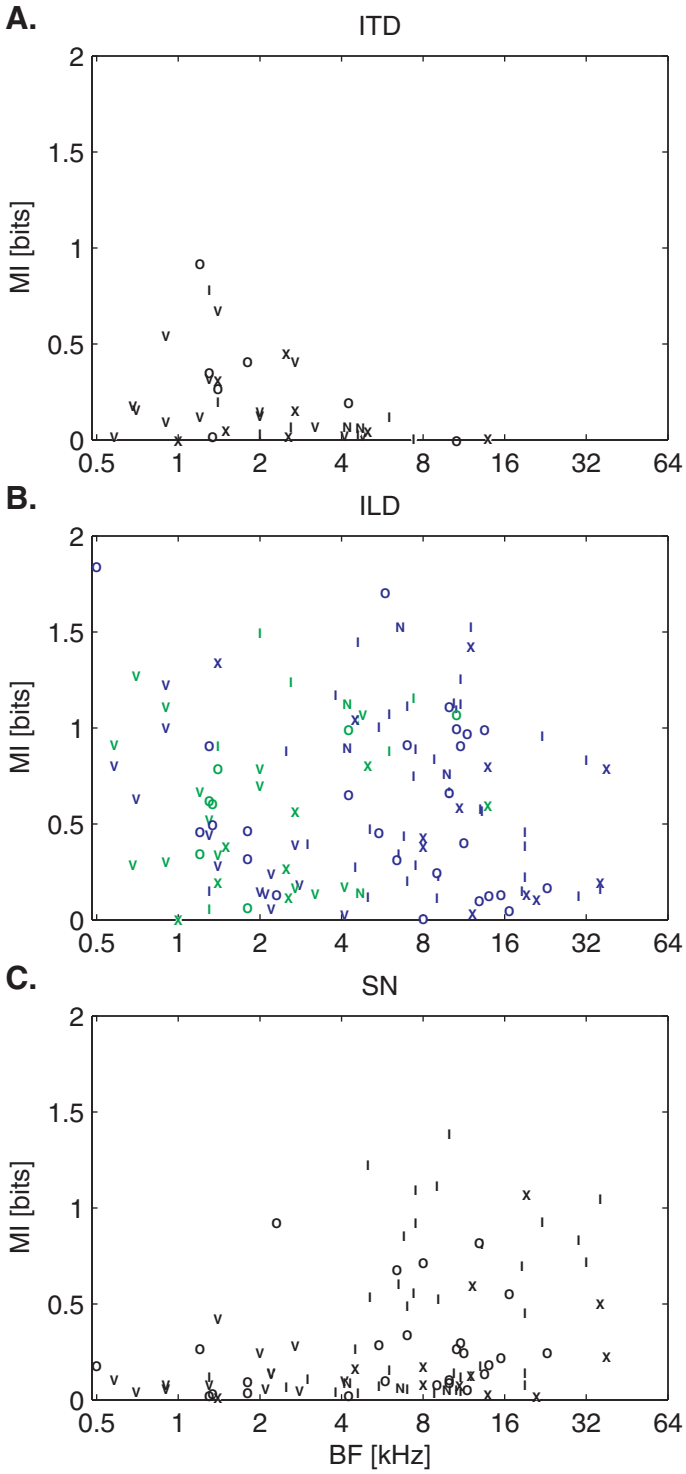


Figure 4.7: Rate information to each individual localization cue for the SN/ILD and ITD/ILD stimulus sets. MI is plotted as a function of each neuron's BF. **A.** ITD. **B.** ILD. *Green:* ITD/ILD stimuli. *Blue:* SN/ILD stimuli. **C.** SN.

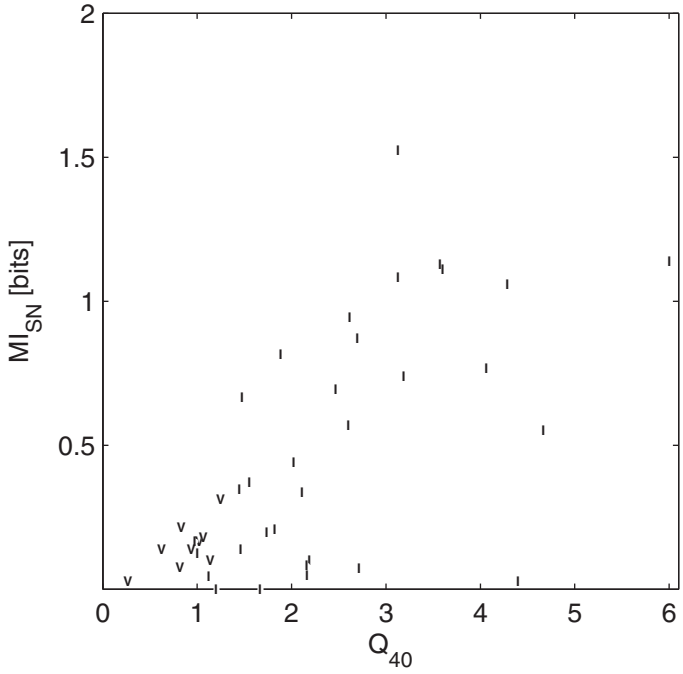


Figure 4.8: MI_{SN} plotted as a function of Q_{40}.

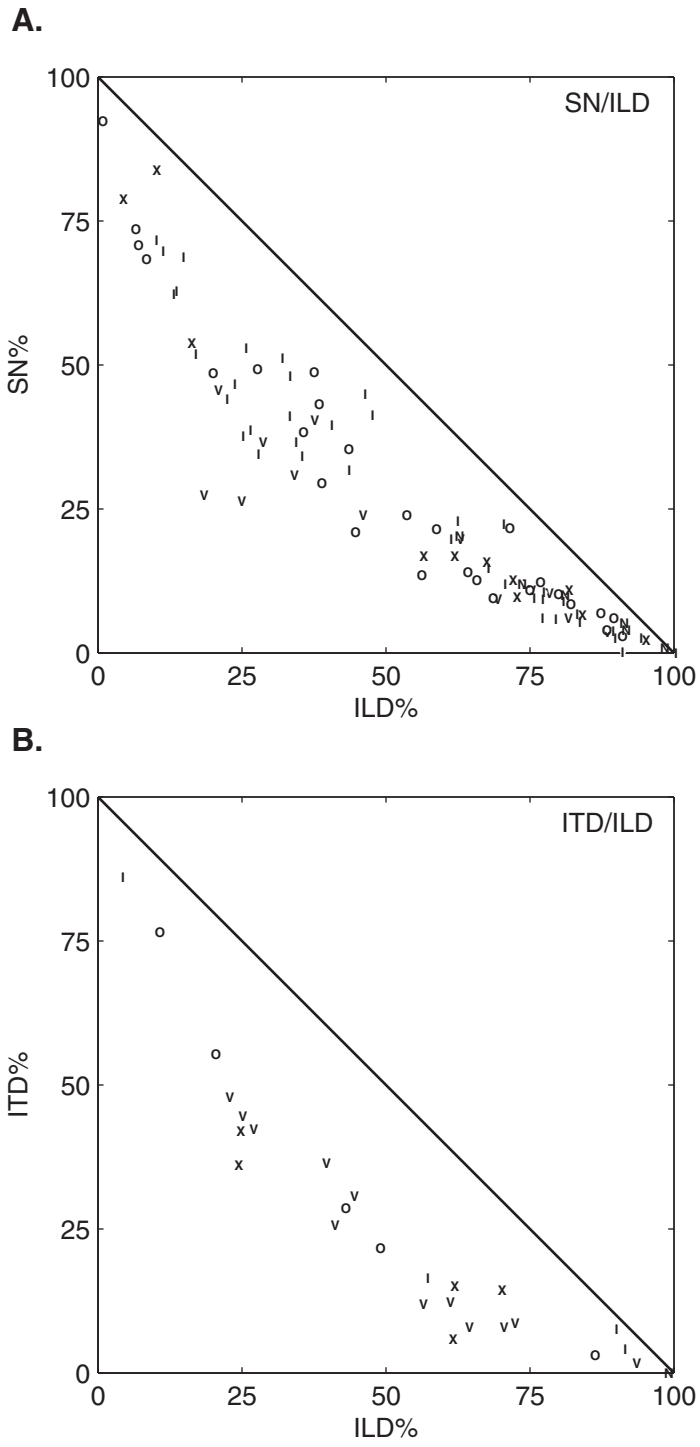


Figure 4.9: The proportion of MI_{FULL} used in coding individual cues. **A.** MI_{SN} is plotted against MI_{ILD} for all neurons studied with the SN/ILD stimulus set, both expressed as a fraction of MI_{FULL}. **B.** Same for MI_{ITD} v. MI_{ILD}. Only neurons with BFs < 4 kHz are plotted in part B.

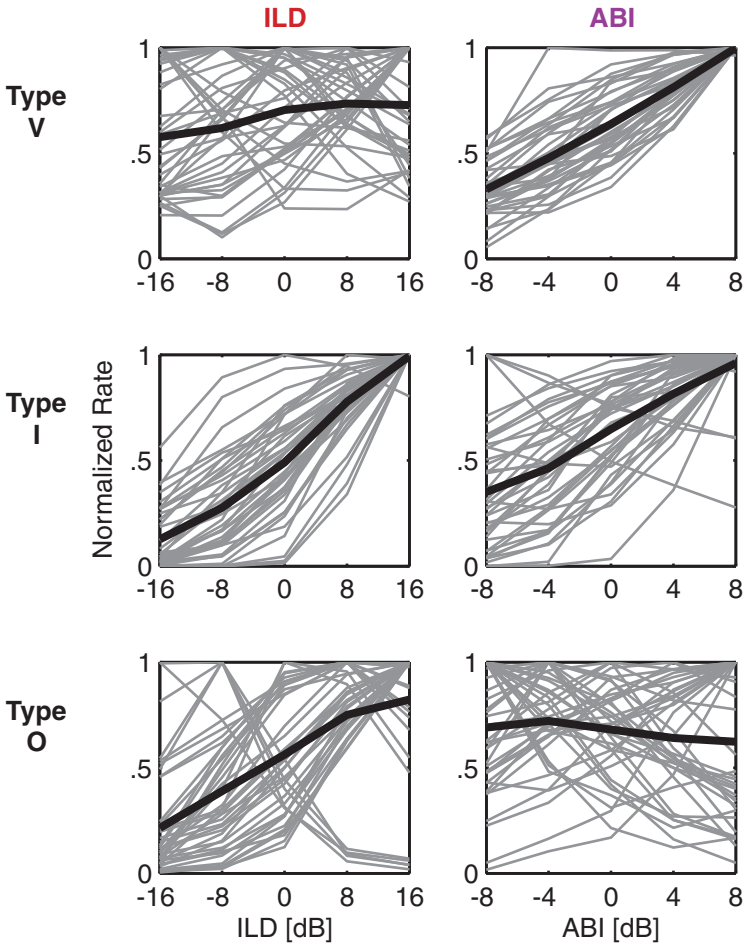


Figure 4.10: Normalized rate profiles of responses to the ABI/ILD stimulus set. Plotting conventions are the same as for Fig. 4.5.

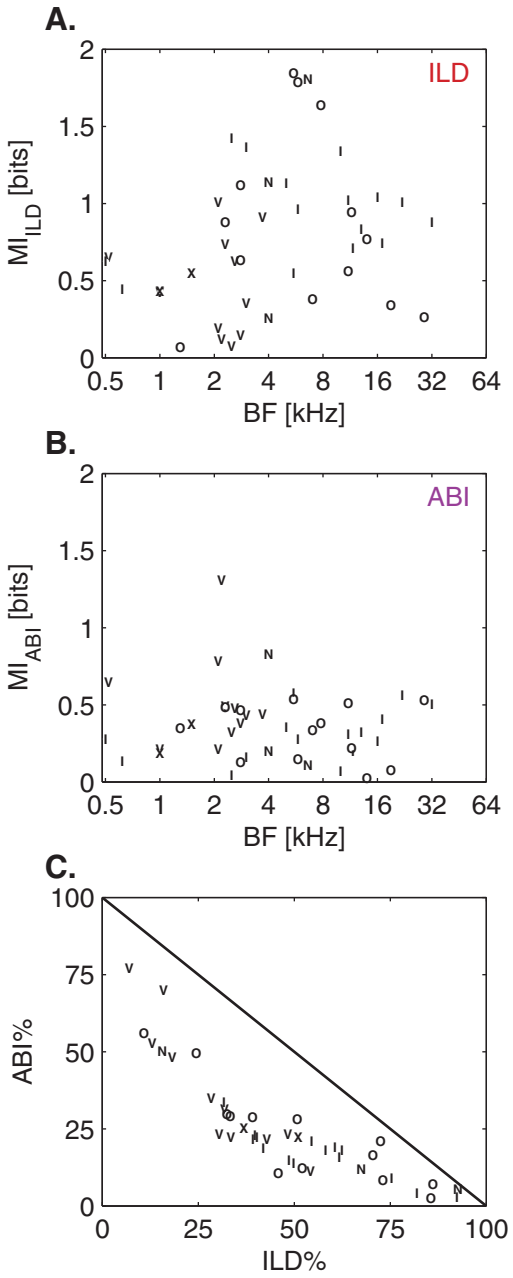


Figure 4.11: The differential coding of average level and level difference information. **A.** MI_{ILD} v. BF for the ABI/ILD stimulus set. **B.** MI_{ABI} v. BF. **C.** MI_{ABI} plotted against MI_{ILD} for each unit, expressed as a percentage of MI_{FULL}.

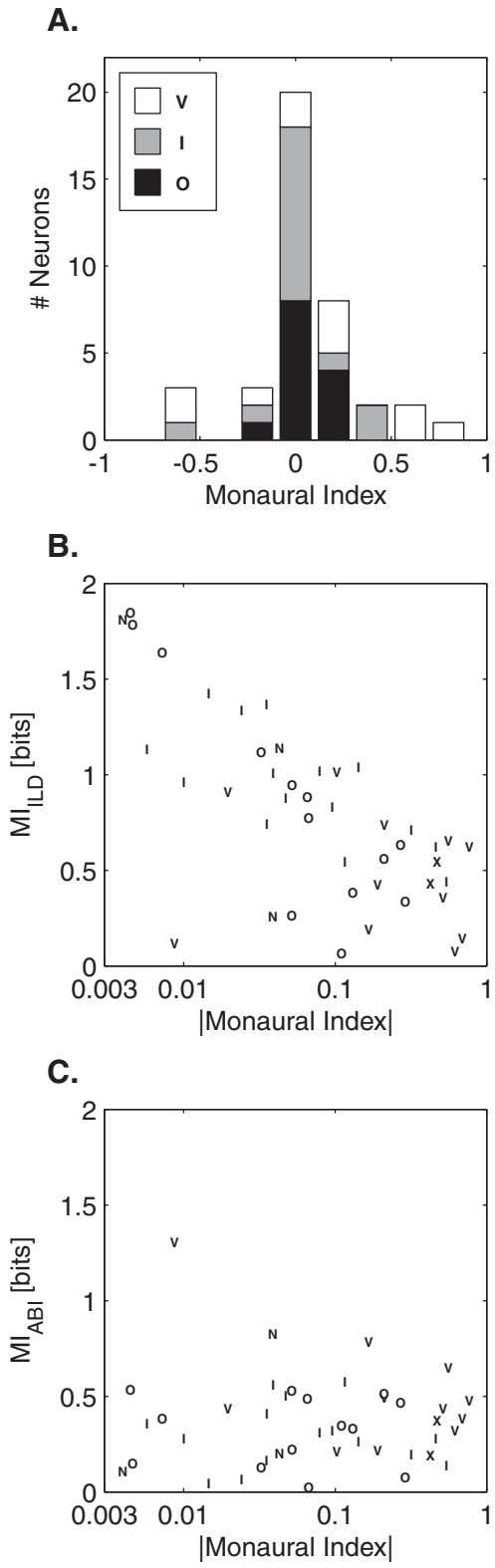


Figure 4.12: Binaural sensitivity and coding. **A.** Histogram of monaural indices broken down by cell type. Values near 0 are the most binaural. Type-O neurons cluster near 0, whereas type-V neurons are more broadly distributed. **B.** MI_{ILD} is plotted against the absolute value of the monaural index, shown on a log scale. **C.** MI_{ABI} v. |monaural index|.

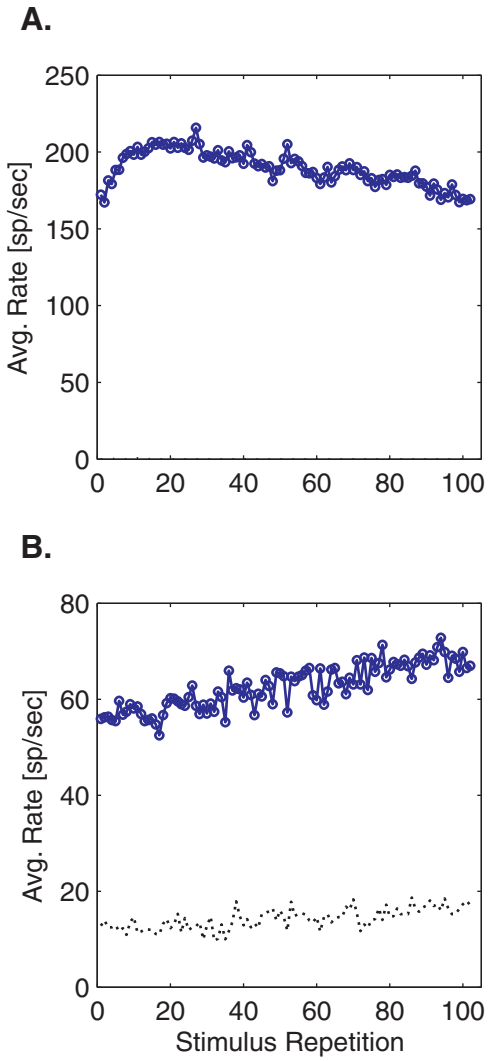


Figure 4.13: Examples of firing rate variance. Each plot shows the average rate (blue circles) to the entire block of 25 stimuli, as a function of the stimulus repetition number. Each plot shows slow drifts in the average firing rate over time. The dashed line gives the average spontaneous rate for the neuron that exhibited spontaneous firing. **A.** 102 repetitions of the SN/ILD stimulus set, presented to a 0.9 kHz type-V neuron. **B.** 102 repetitions of the ITD/ILD stimulus set, presented to a 1.2 kHz type-O neuron.

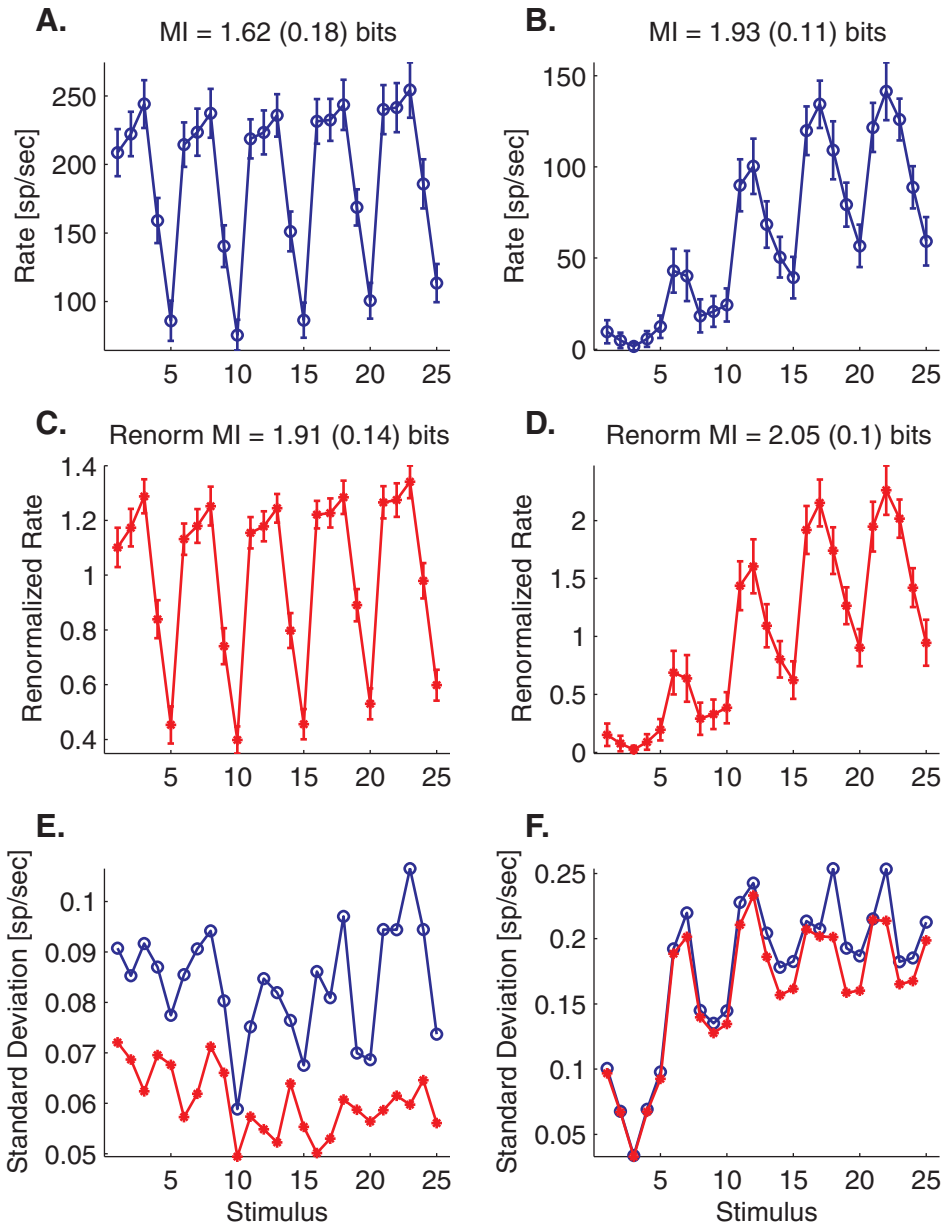


Figure 4.14: Renormalized MI examples. **A.** Mean and standard deviation of the firing rates to the 25 SN/ILD stimuli of the neuron in Fig. 4.13A. The full MI is given in the title bar, with the bias value in parentheses. **C.** Mean and standard deviation of the renormalized firing rates of the neuorn in A. The y-axis limits have been chosen to show the same relative scale as the firing rates in A, such that the relative magnitudes of the error bars can be compared. The MI calculated from the renormalized firing rates is shown in the title. **E.** *Blue circles:* Standard deviations of the responses in A, normalized by the average firing rate across all trials and all stimuli. These are the expected standard deviations of the normalized firing rates in B, under the assumption of a stationary process. *Red asterisks:* Actual standard deviations of the normalized firing rates in B. **B, D, F.** Same as A, C, and E, for the example neuron of Fig. 4.13B.

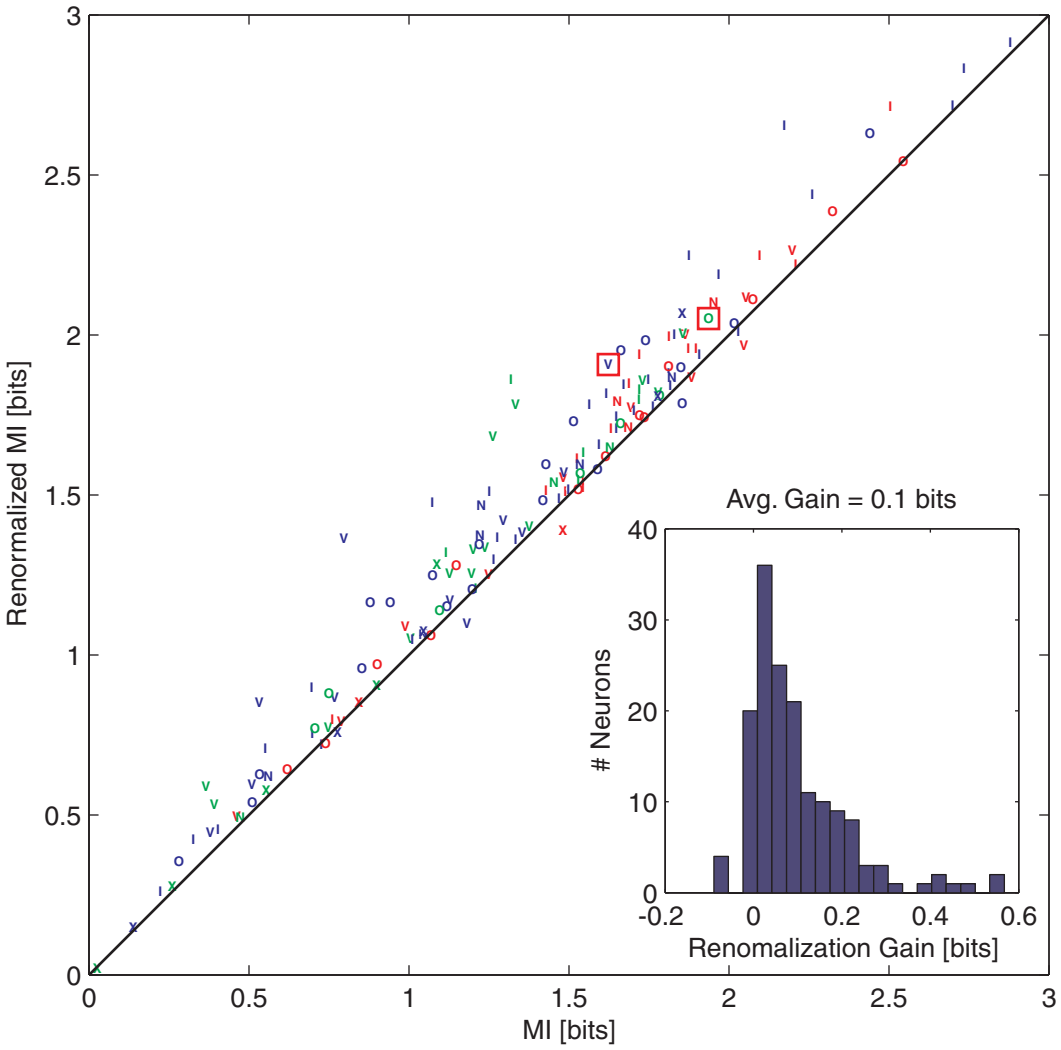


Figure 4.15: Population renormalization gain. The renormalized MI is plotted as a function of the MI calculated directly from spike count for all neurons. The example neurons of Figs. 4.13 and 4.14 are outlined in red. Colors denote the stimulus set, with green for ITD/ILD stimuli, blue for SN/ILD stimuli, and red for ABI/ILD stimuli. *Inset:* histogram of the renormalization gains, calculated as the difference between the renormalized MI and the normal MI. The average gain is 0.1 bits.

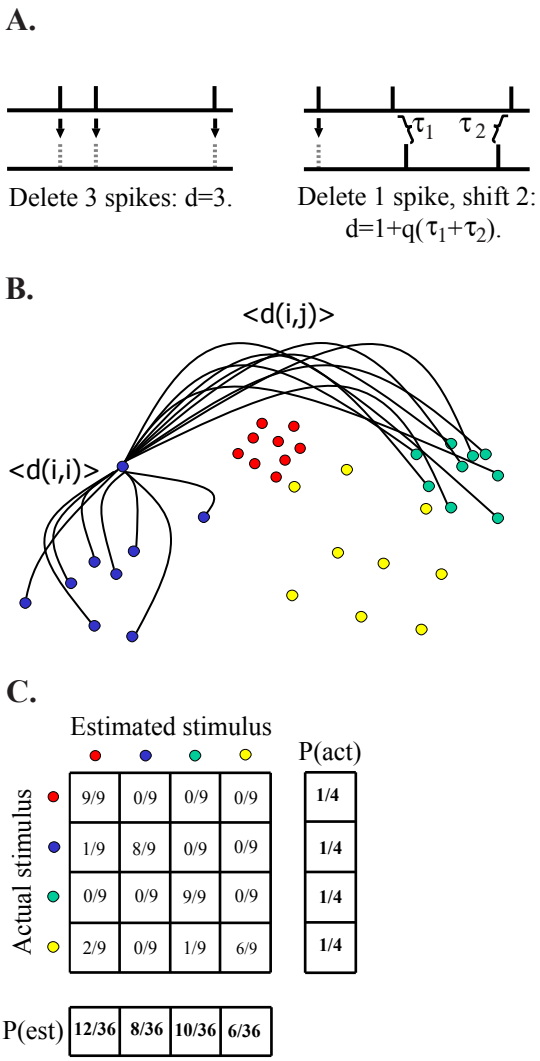


Figure 5.1: MI calculation using the spike distance metric. **A.** Spike distance calculation examples. *Left:* To turn the top spike train into the bottom spike train, three spikes must be deleted. *Right:* To turn the top spike train into the bottom spike train, one spike must be deleted, and two spikes shifted. This is the minimum-cost solution for all cost parameters q such that $q|\tau_i|<2$ for all i. When this condition is not met, the spike is deleted from its position in the top train and added to the corresponding spot in the bottom train, adding a distance of 2. **B.** Stimulus clustering with the SDM method. Each dot represents a spike train whose color denotes the stimulus played when the train was recorded. For each train, the average distance $\langle d(i,j) \rangle$ to each group of spike trains is calculated. The spike train is estimated to have come from the stimulus group to which the average distance is smallest. **C.** A confusion matrix counts the number of spike trains assigned to each stimulus class. When normalized by the total number of spike trains, this is the joint stimulus response matrix used to calculate the MI.

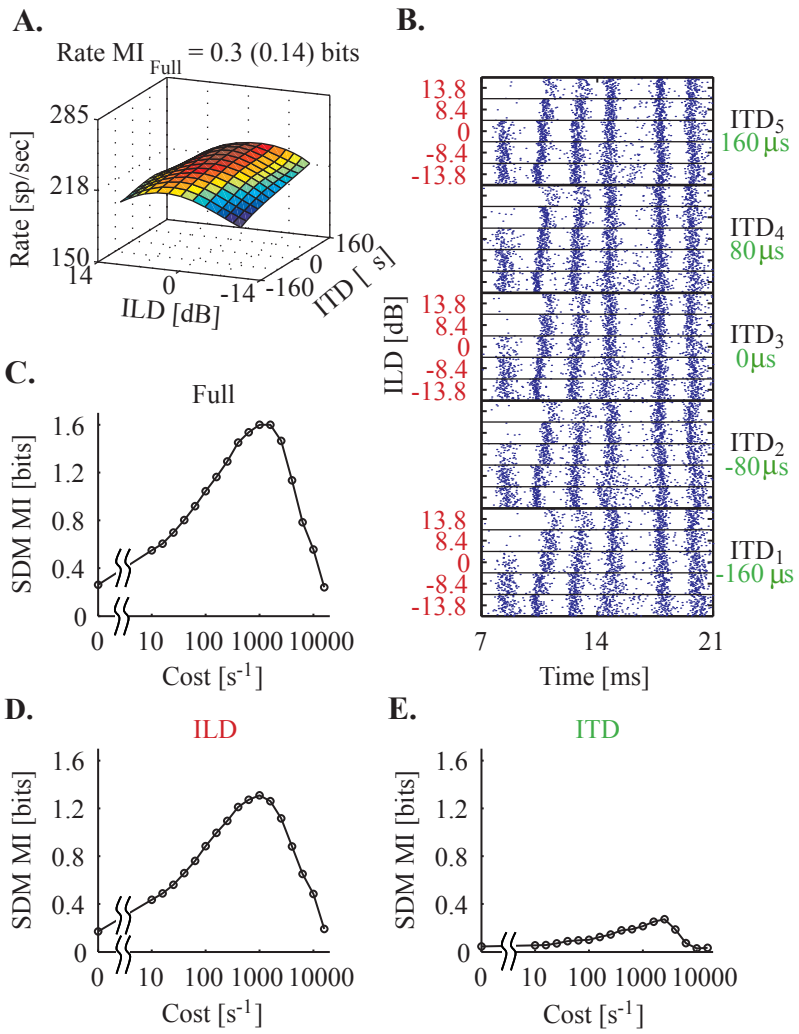


Figure 5.2: Responses of a 4.8 kHz type-V neuron to the ITD/ILD stimulus set. **A.** Surface plot of the mean spike rate as a function of ITD and ILD. The rMI is shown above the plot, with the bias value in parentheses. (The surface was smoothed with cubic-spline interpolation.) **B.** Spike rasters to all 25 stimuli showing the first 6 spike bursts. Thick black lines divide different ITD parameters, shown at right; thin black lines divide different ILD parameters, shown at left. 110 repetitions of each stimulus were collected. **C.** MI measured with the SDM method as a function of the spike shift cost parameter for the full stimulus set. Note that zero cost has been put at the 1 s^{-1} position. **D.** MI_{ILD} v. cost. **E.** MI_{ITD} v. cost.

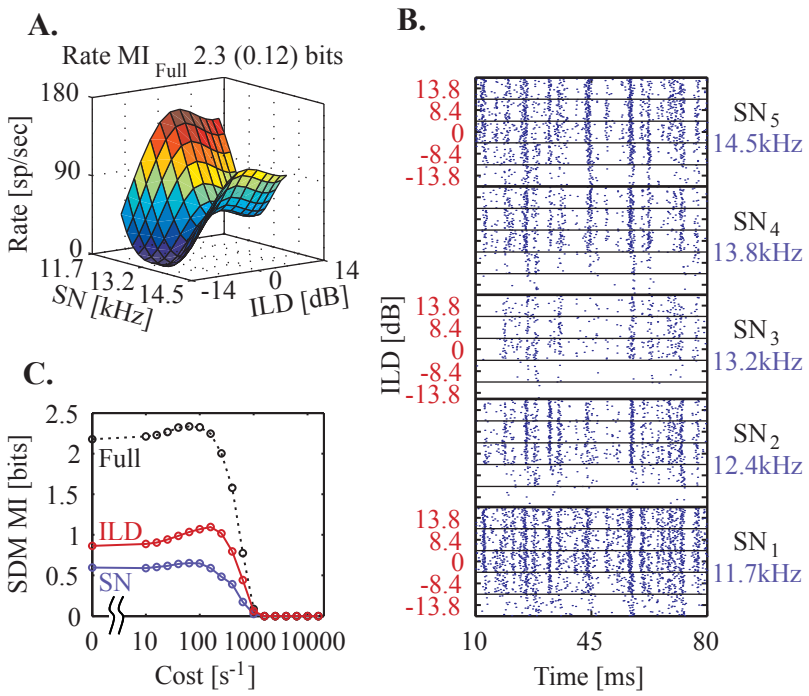


Figure 5.3: Another example of information extraction with the SDM method, this time for a 13.2 kHz type-I neuron. The format is as in Fig. 5.2, with the exception that the three MI(cost) curves have been condensed to a single plot.

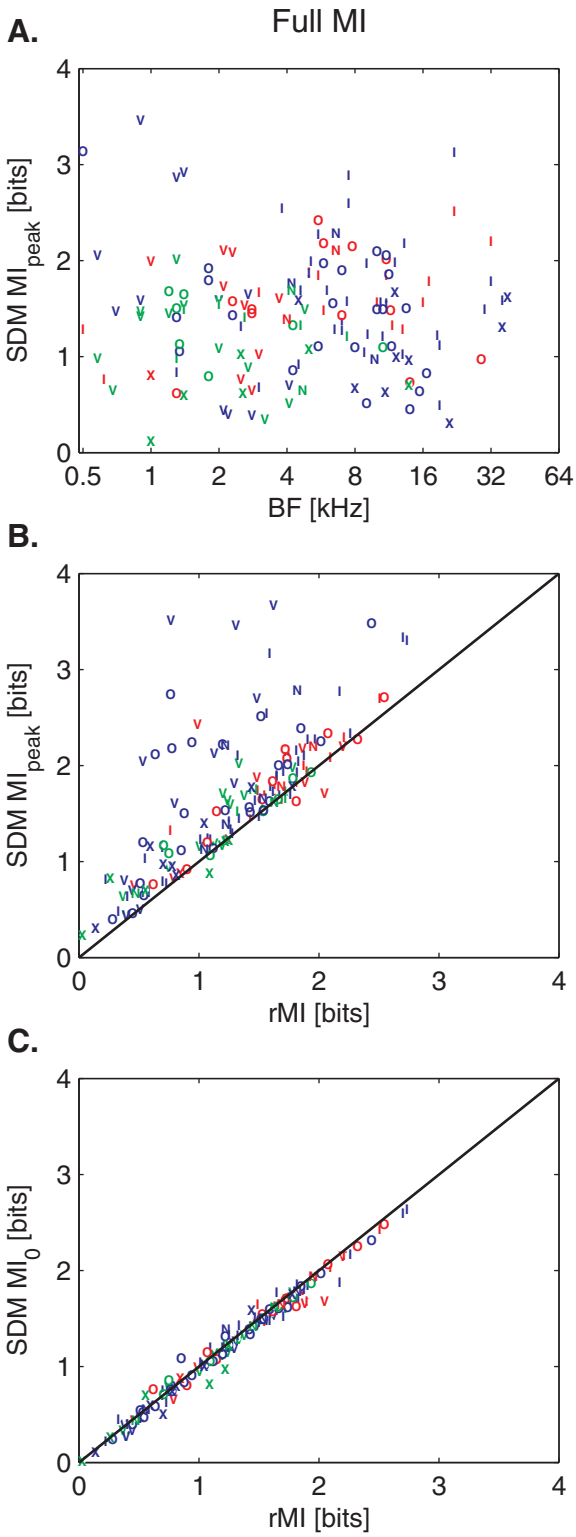


Figure 5.4: SDM information carried about the full stimulus set. **A.** MI_{peak} plotted as a function of BF for all neurons and stimuli. **B.** MI_{peak} as a function of the information calculated directly from the spike rates. **C.** MI from the SDM method at zero cost plotted as a function of the MI calculated directly from discharge rate. Colors denote the stimulus set, with green for ITD/ILD, blue for SN/ILD and green for ITD/ILD.

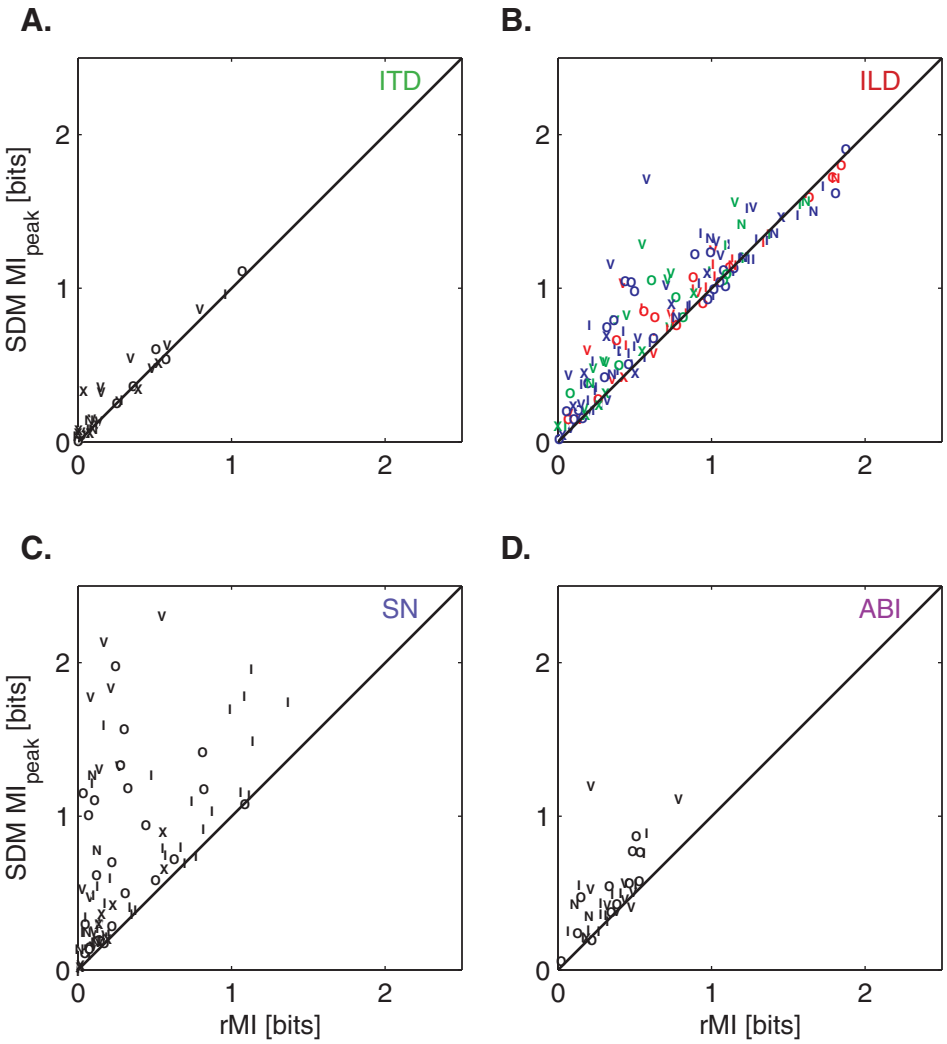


Figure 5.5: Comparison between the peak information calculated with the SDM method and rMI for individual localization cues. **A.** ITD. **B.** ILD. Colors denote the stimulus set, with green for ITD/ILD, blue for SN/ILD, and red for ABI/ILD. **C.** SN. **D.** ABI.

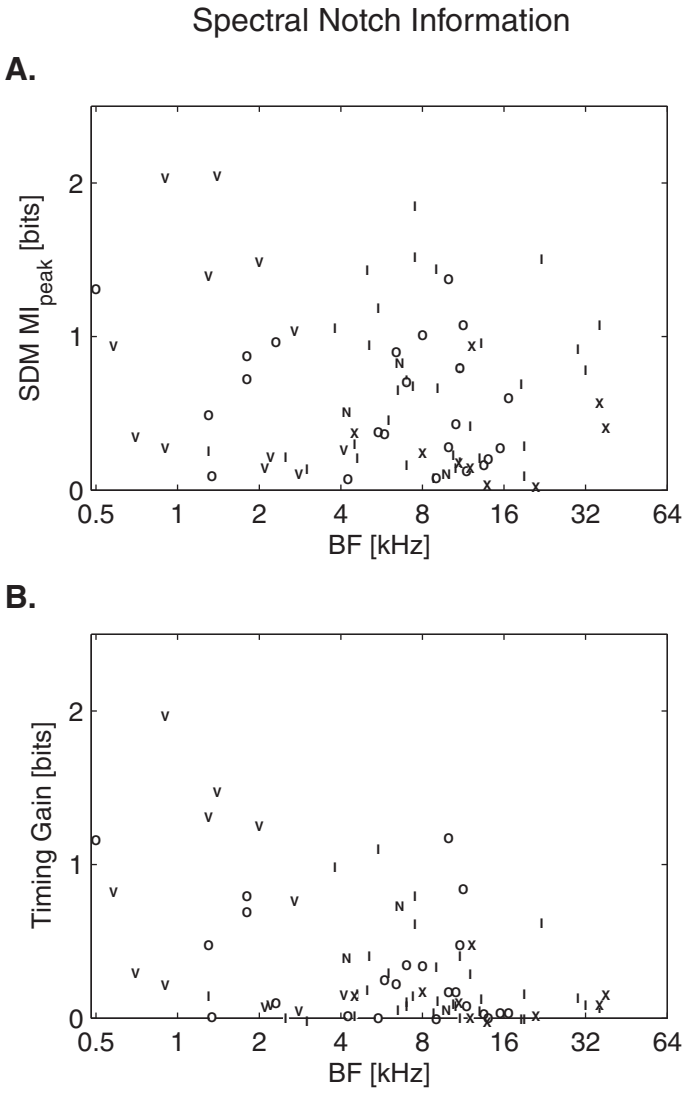


Figure 5.6: SN timing gain. **A.** Peak SDM information plotted as a function of MI for the SN cue. **B.** SN spike timing gain (MI_{peak}-MI_0) as a function of BF.

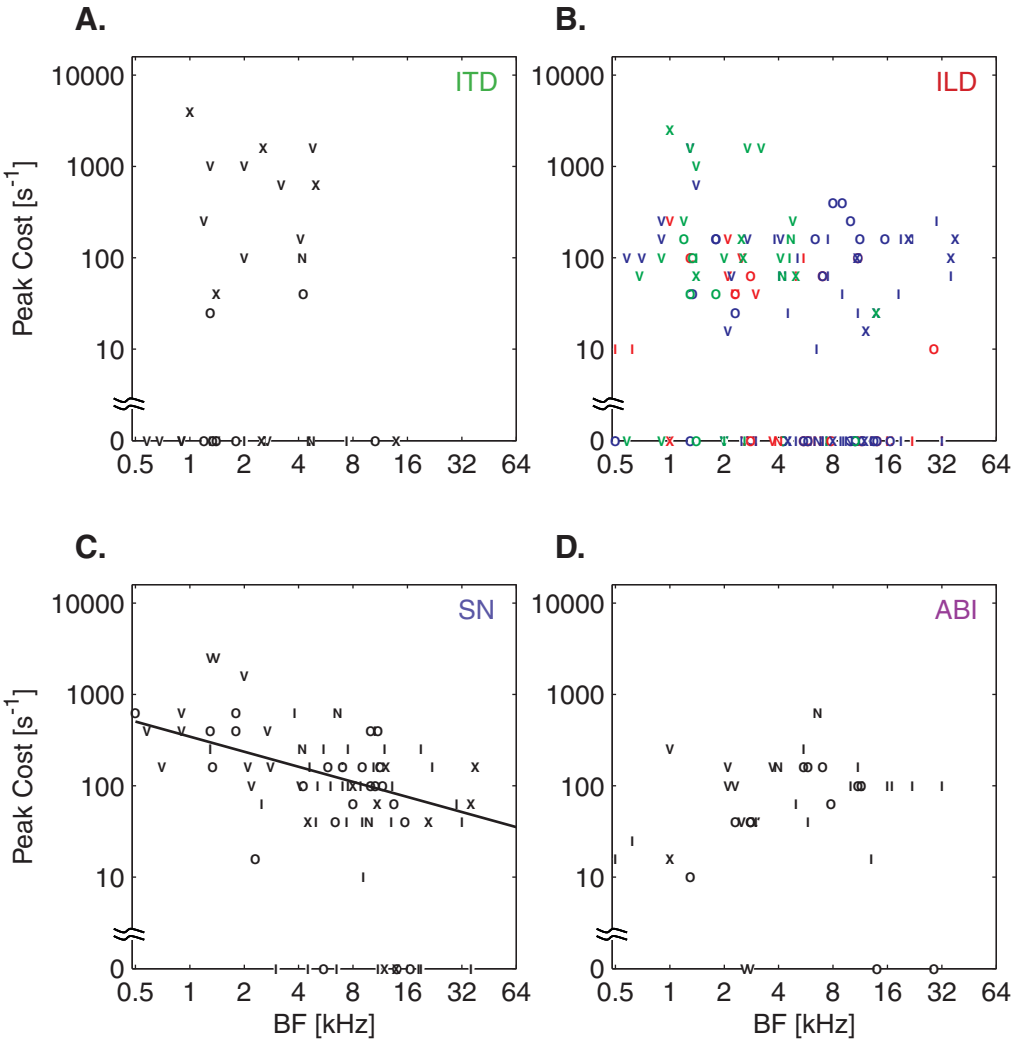


Figure 5.7: Time scale of the temporal representation of localization cues. **A.** Cost at which the SDM MI reached its peak value plotted as a function of BF for the ITD cue. Note that the zero-cost data points have been placed at the $1s^{-1}$ spot. **B.** ILD. **C.** ABI. **D.** SN. The black line represents the best linear fit of $\log_{10}$(cost) to $\log_2$(BF), ignoring zero cost values.

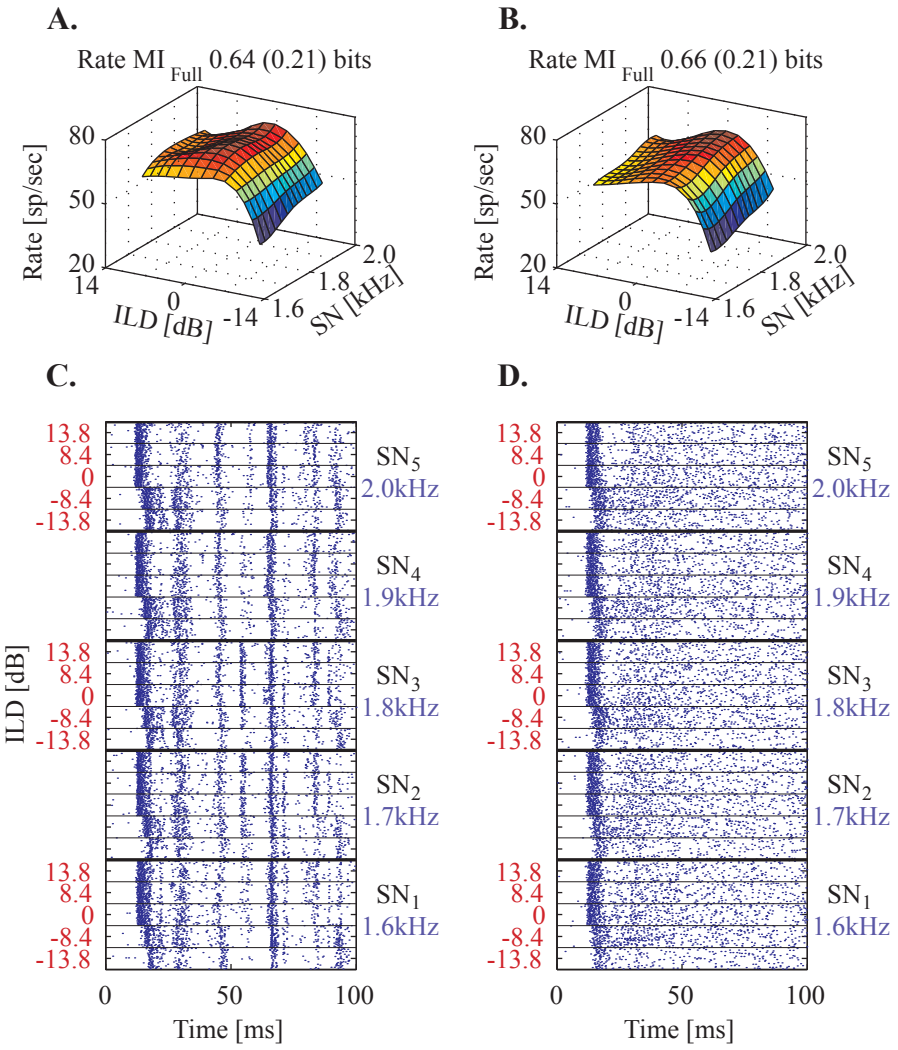


Figure 5.8: Responses of a 1.8 kHz type-O neuron to the frozen and random noise stimulus sets. **A.** Average firing rate as a function of SN and ILD for the frozen noise stimulus set. **B.** Same, for the random noise set. **C.** Spike rasters of the responses to the frozen noise stimulus set. **D.** Rasters of the responses to the random noise set. Note that the ILD-dependent latency differences remain, while most other time-locked responses vanish.

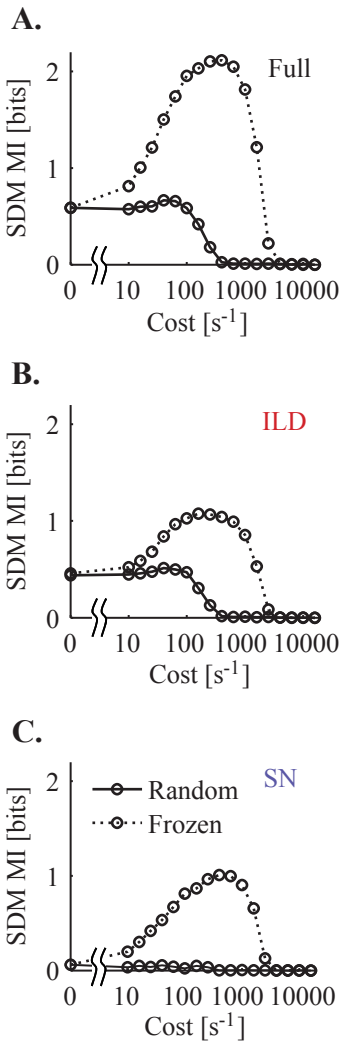


Figure 5.9: SDM coding of random and frozen noise stimuli. This example neuron is the same neuron as in Fig. 5.8. **A.** SDM MI_{FULL} as a function of cost for the frozen (dotted) and random (solid) stimulus sets. **B.** SDM MI_{ILD}(cost). **C.** SDM MI_{SN}(cost).

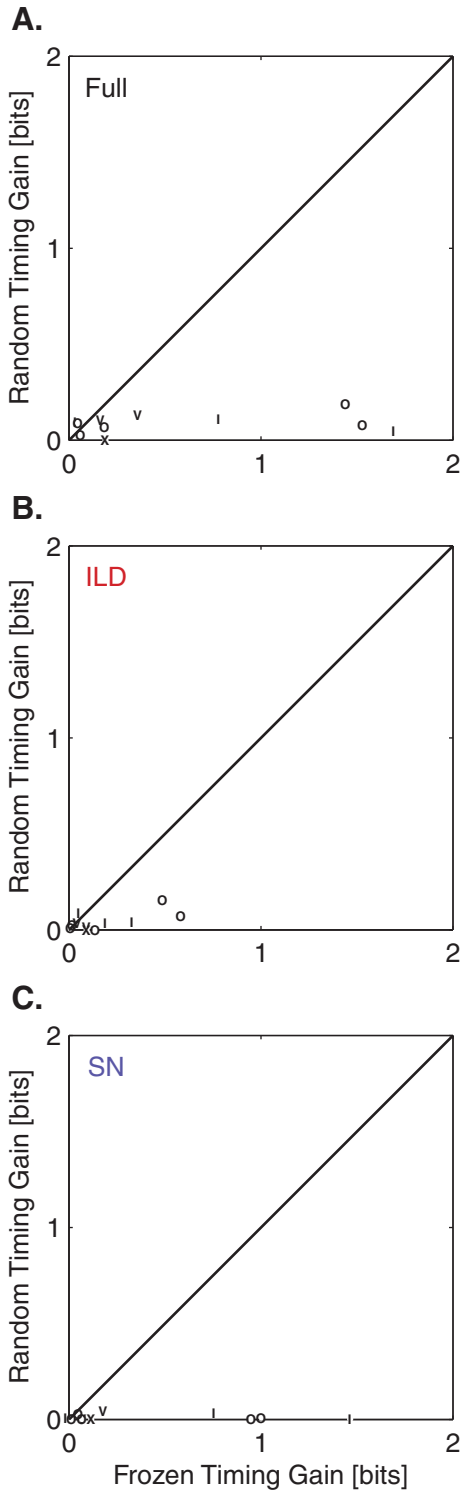


Figure 5.10: Comparison of spike timing gains for frozen and random noise stimuli. In each plot, the spike timing gain to the random noise stimulus set is plotted as a function of the spike timing gain to the frozen set. **A.** Gain comparison for MI_{FULL}. **B.** MI_{ILD}. **C.** MI_{SN}.

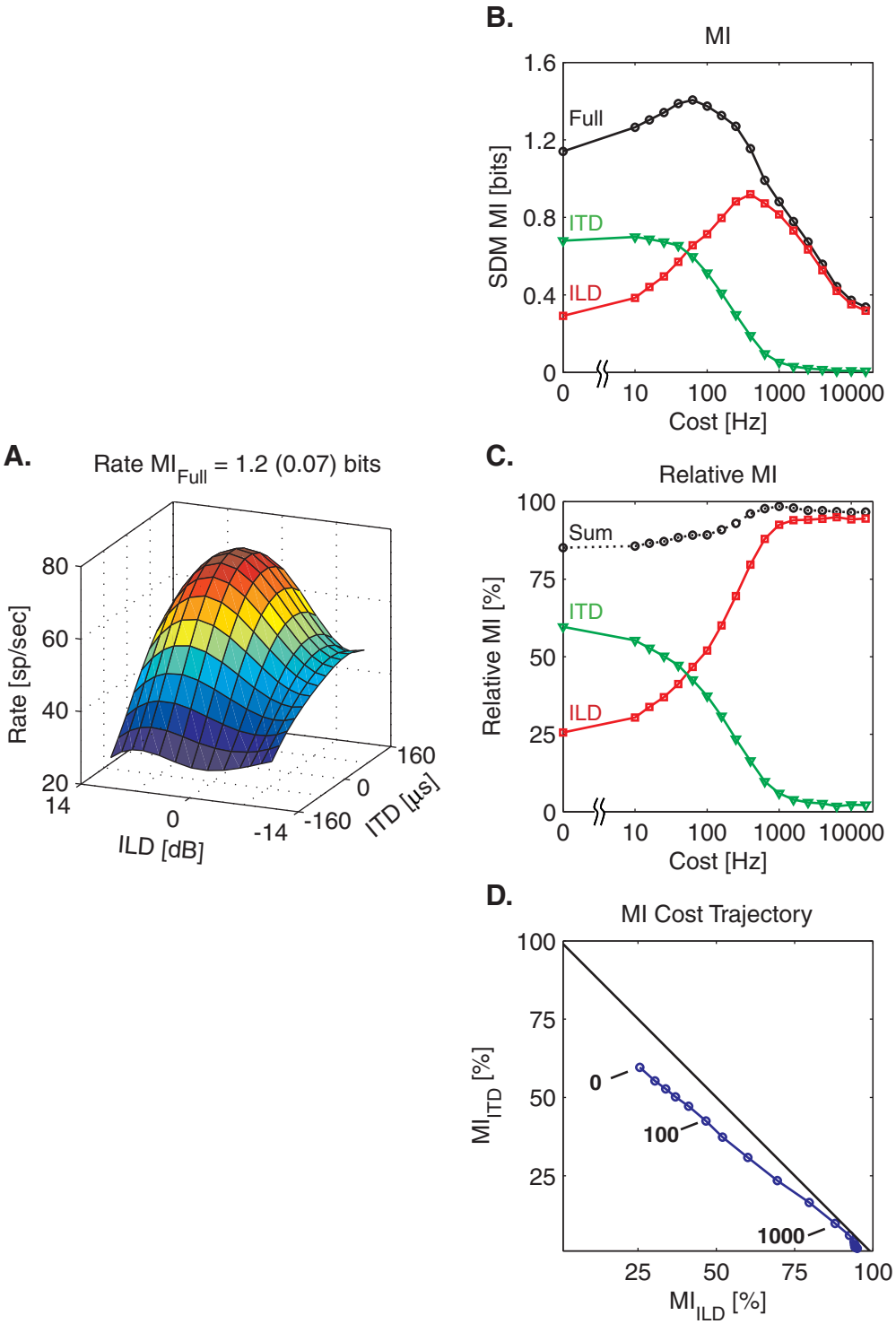


Figure 5.11: The temporal coding of multiple cues. **A.** The average rate response of a 1.4 kHz type-V neuron in response to the ITD/ILD stimulus set. **B.** MI(cost) curves for this neuron. **C.** Relative information (MI [%]) plotted as a function of cost. Their sum is plotted as a dotted line. **D.** The percentage of information devoted to ITD coding plotted as a function of that devoted to ILD coding for each cost studied; certain trajectory points are labeled with their cost values [s^{-1}].

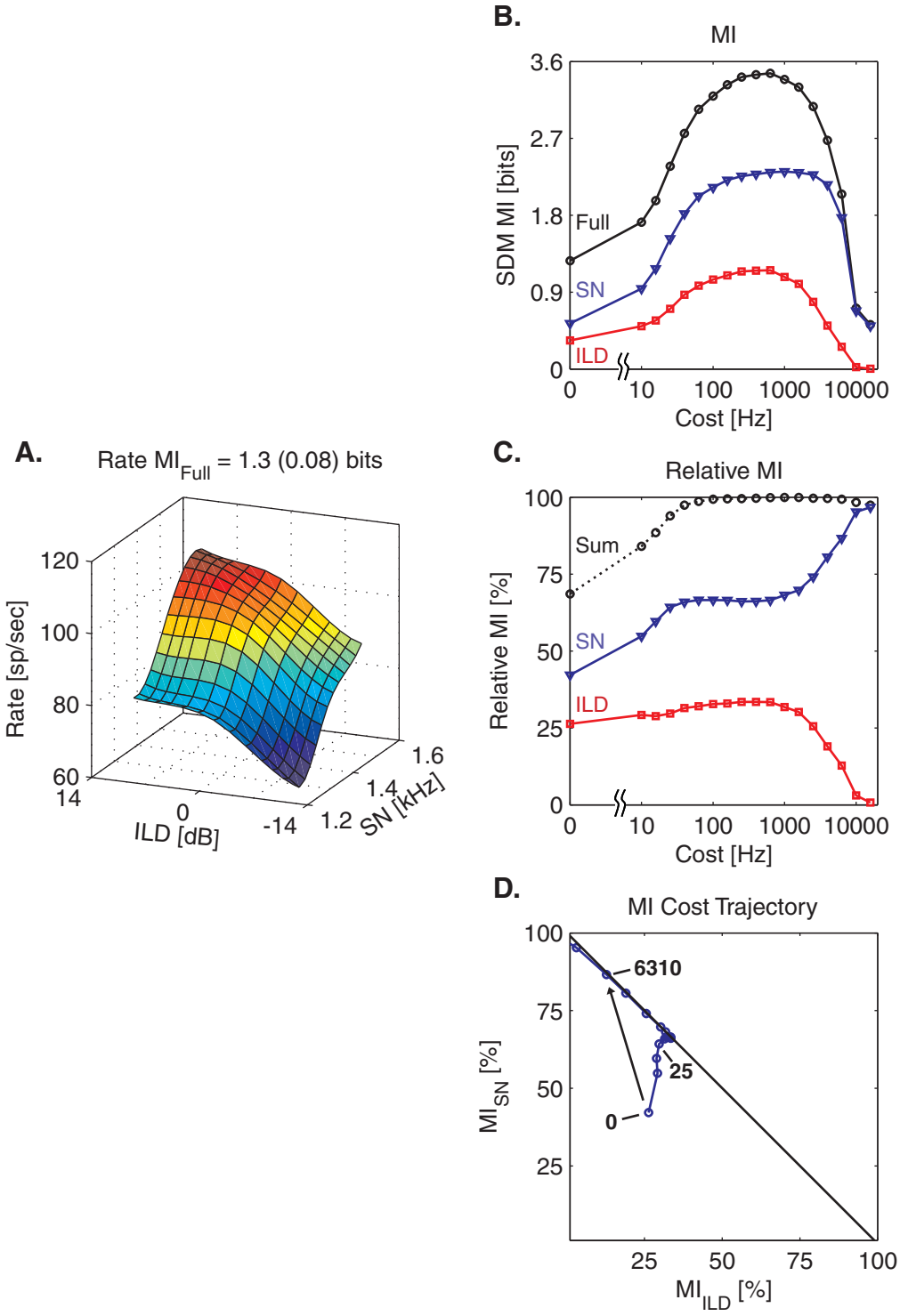


Figure 5.12: Building a cost/coding trajectory. This is another example of the temporal coding of multiple cues, from the same neuron as Fig. 5.11, but this time to the SN/ILD stimulus set. **A.** Mean rate responses. **B.** MI(cost) curves. **C.** Relative information as a function of cost. **D.** Cost trajectory for SN/ILD cues. The arrow represents the summary trajectory from zero cost to the cutoff cost (6310 s^{-1} for this neuron).

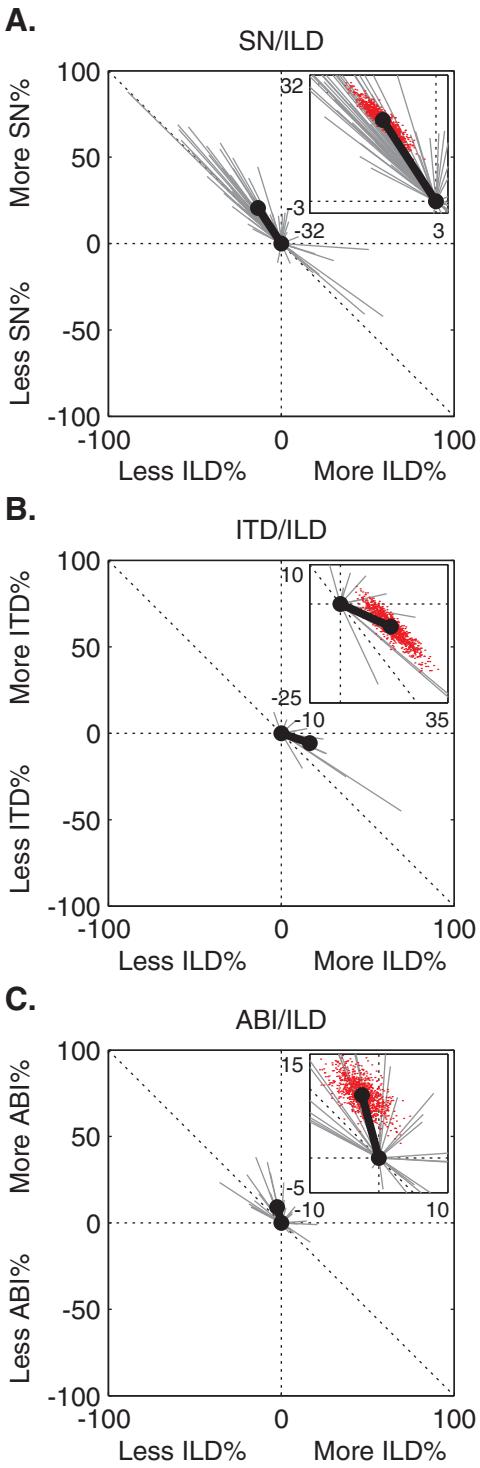


Figure 5.13: Population coding trajectories. Each grey line represents the change in relative MI from zero cost (origin) to the cutoff cost for that neuron. The population mean is given by the thick black line. *Inset:* close-up of the mean trajectory vector, where the red dots represent the mean trajectory vector endpoints for 1000 bootstrap resamplings of the underlying individual trajectories. **A.** SN/ILD stimuli. **B.** ITD/ILD stimuli. **C.** ABI/ILD stimuli.

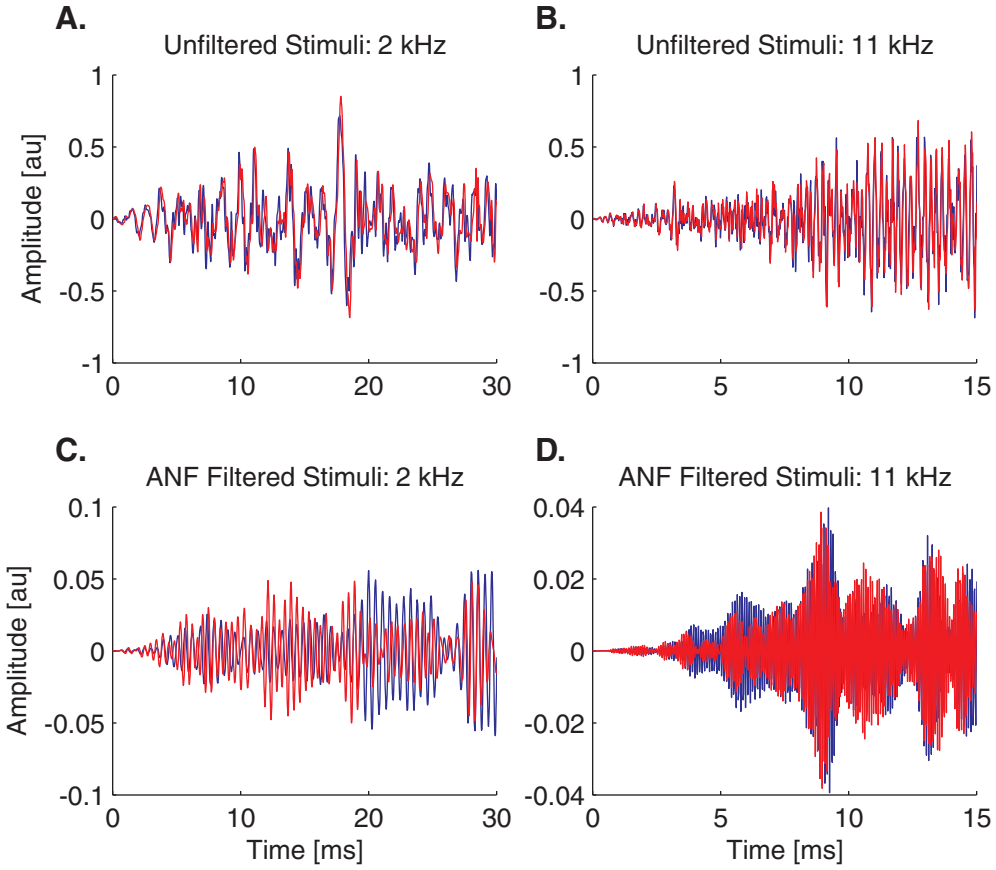
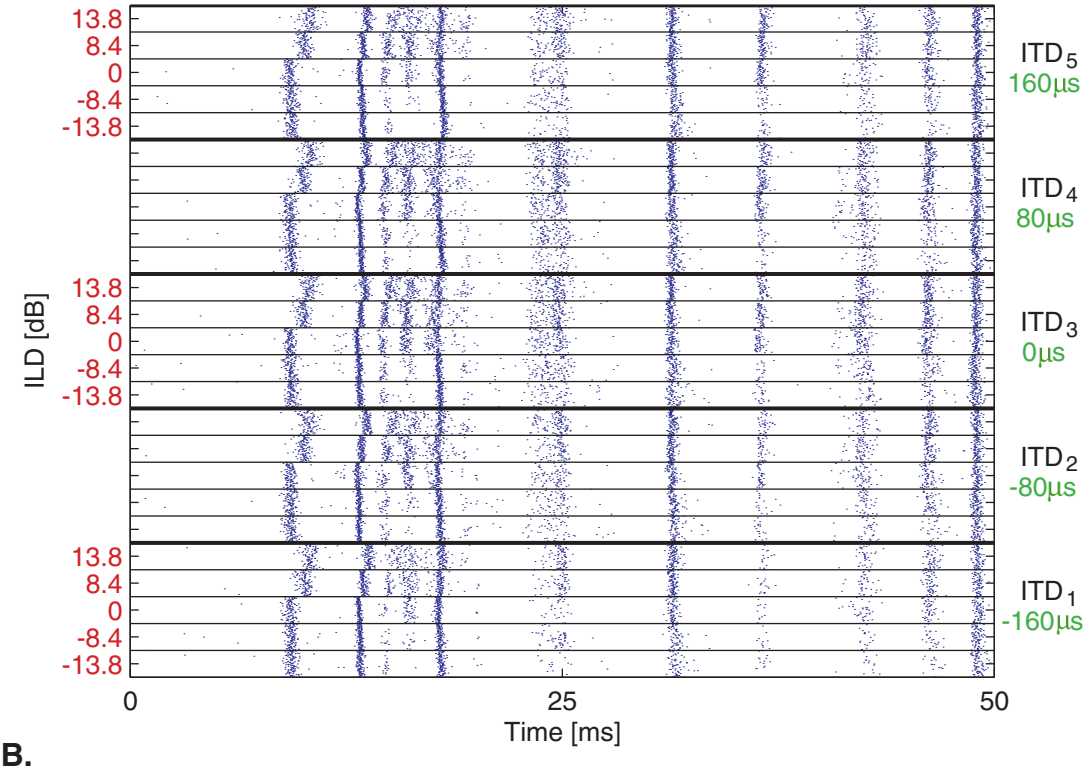


Figure 5.14: Example temporal waveforms for SN/ILD stimuli resampled to shift the center SN (red) to 2 kHz (A, C) or 11 kHz (B, D). The blue waveform shows the first SN, which is approximately 0.2 octaves below the center notch. **A, B.** Raw stimulus waveforms, after resampling. **C, D.** The stimulus waveforms shown in A and B, after filtering that mimics an auditory nerve fiber with a BF of 2 kHz (C) or 11 kHz (D).

A.


13.8
8.4
0
-8.4
-13.8
ITD5
160µs
ITD4
80µs
ITD3
0µs
ITD2
-80µs
ITD1
-160µs
ILD [dB]
0
25
50
Time [ms]

B.

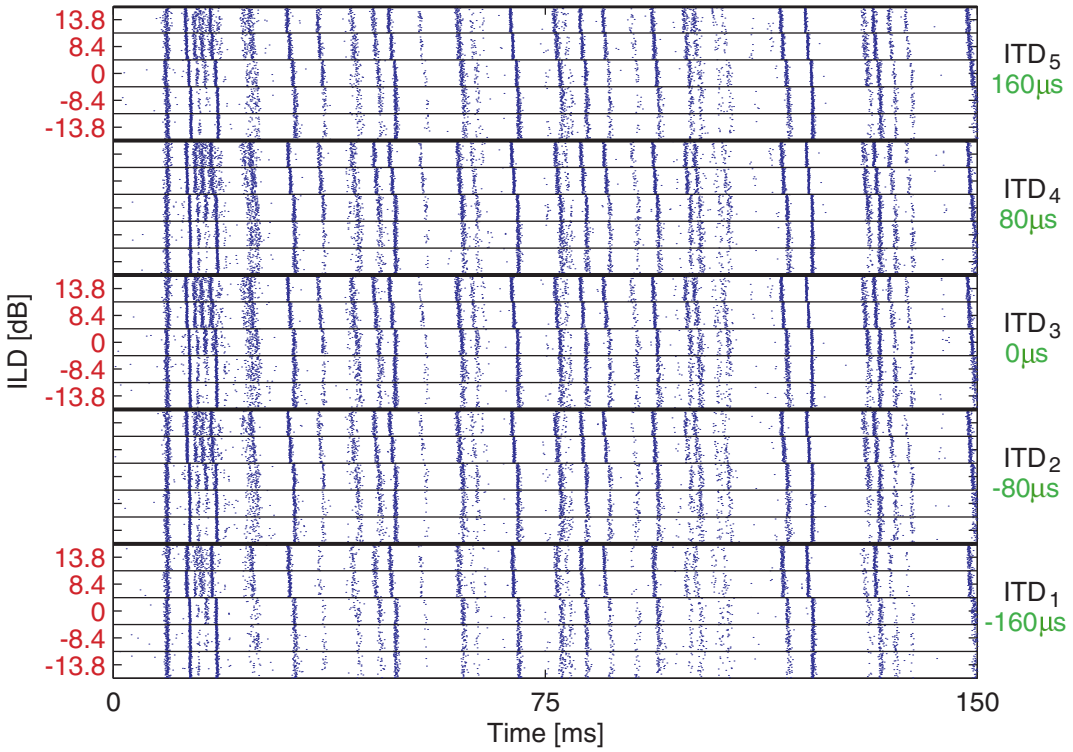

13.8
8.4
0
-8.4
-13.8
ITD5
160µs
ITD4
80µs
ITD3
0µs
ITD2
-80µs
ITD1
-160µs
ILD [dB]
0
75
150
Time [ms]

Figure 6.1: First spikes are special. **A.** Spike raster plot of the first 50 ms of one neuron's response to the ITD/ILD stimulus set. **B.** Same rasters as in A, but 150 ms of response are shown and the spike trains have been shifted such that the first spike bursts overlap (see text for details).

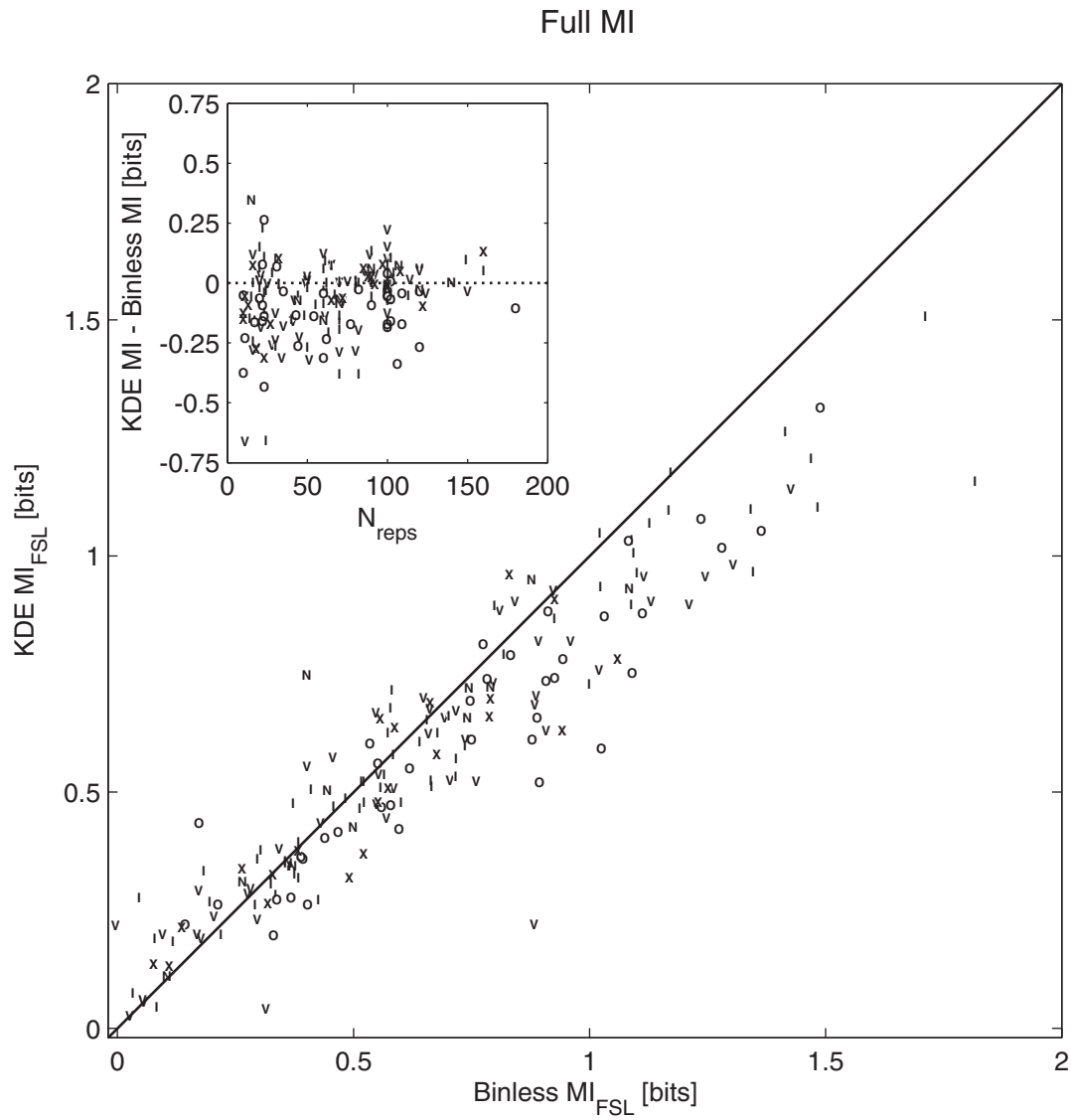


Figure 6.2: Comparison of kernel density and binless estimates of MI_{FSL}, for MI estimates that have not been bias corrected. The full MI estimated with the KDE method is plotted as a function of that estimated with the binless method for all neurons. *Inset:* the difference between the KDE and binless estimates is plotted as a function of the number of stimulus repetitions.

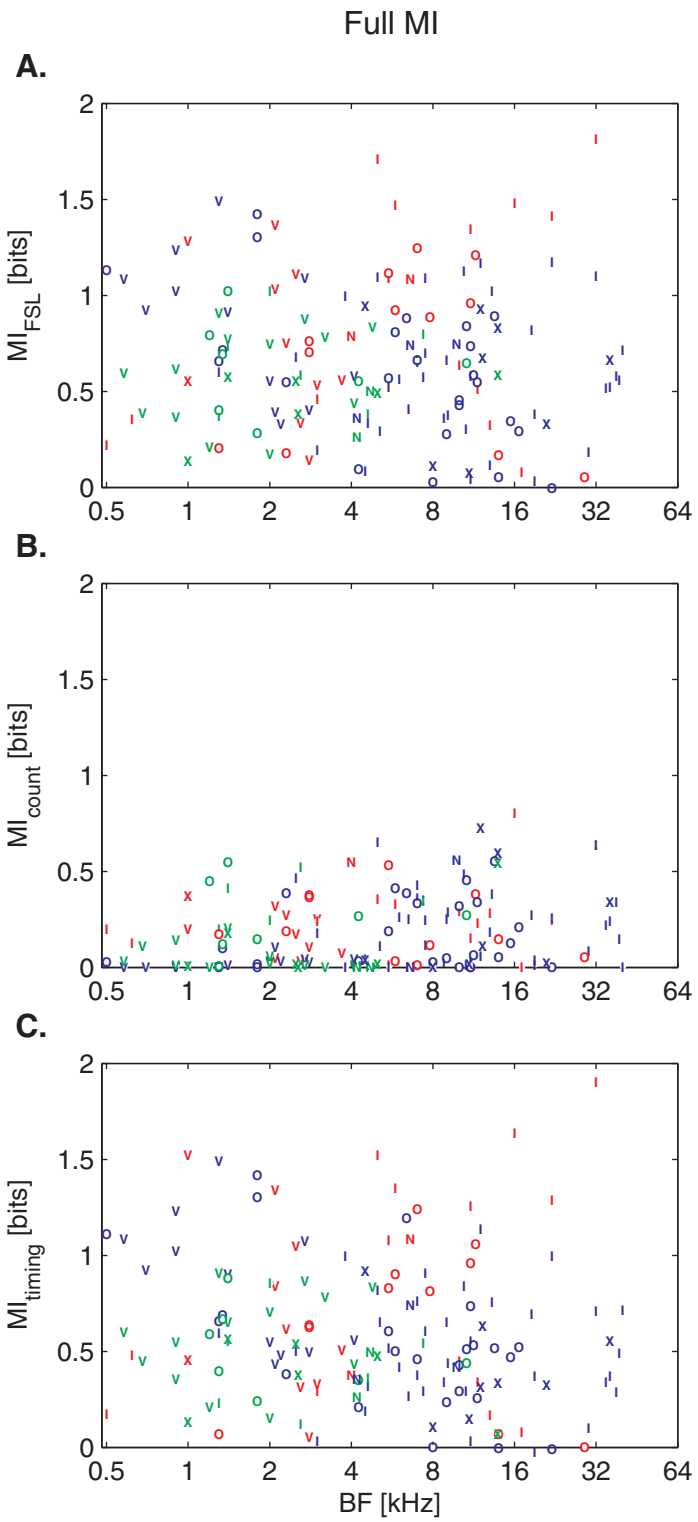


Figure 6.3: MI_{FSL} for the full stimulus set. Colors denote the stimulus set, with green for ITD/ILD, blue for SN/ILD, and green for ABI/ILD. **A.** MI_{FSL}, computed with the binless method, plotted as a function of BF. **B.** MI_{count} as a function of BF. **C.** MI_{timing} as a function of BF. Two points with negative MI_{timing} have been removed.

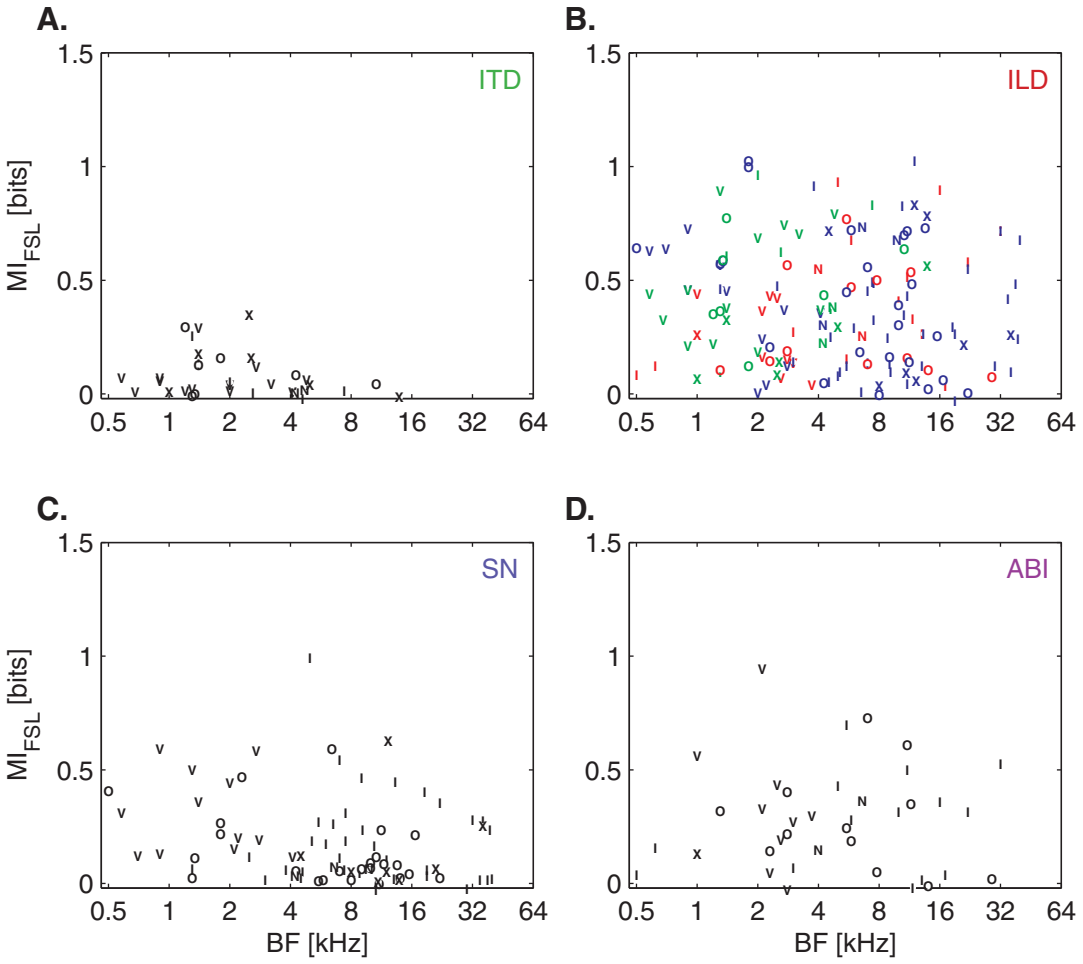


Figure 6.4: First spike latency differentially encodes different localization cues. **A.** MI_{FSL} for the ITD cue is plotted as a function of BF. **B.** Same, for ILD. Colors denote the stimulus sets, as in Fig. 6.3. **C.** SN. One point with negative MI_{FSL} was removed. **D.** ABI.

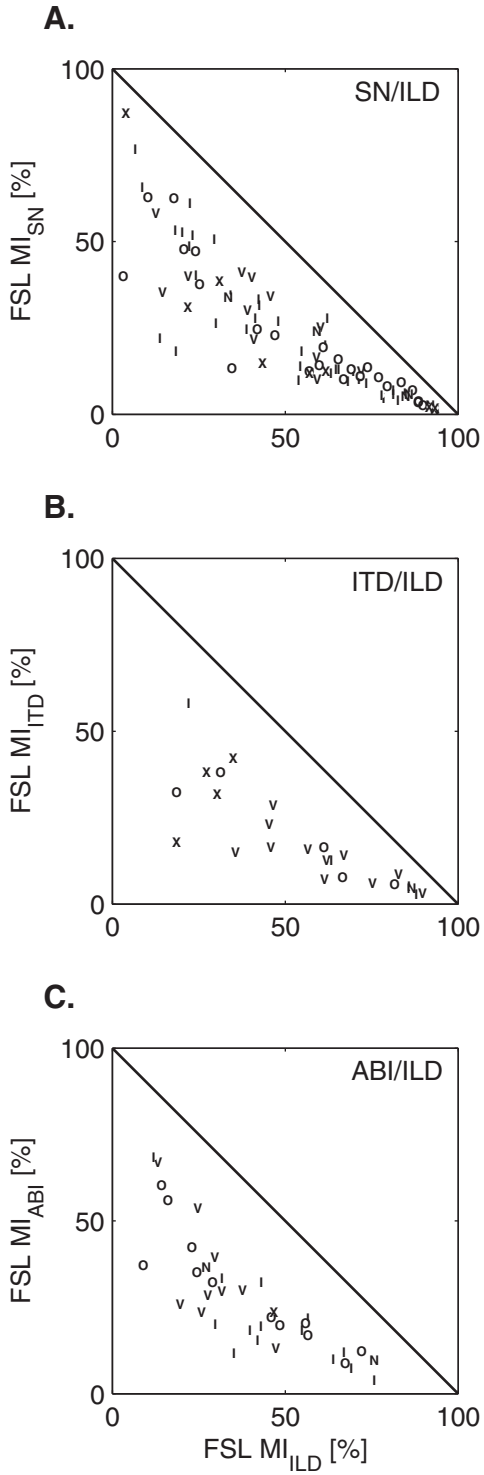


Figure 6.5: First spike latency coding of multiple cues. The relative MI for one cue in each set is plotted as a function of the relative MI for the other cue. (The KDE method was used for these MI estimates). **A.** SN/ILD. **B.** ITD/ILD. Only neurons with BFs < 4 kHz are included. **C.** ABI/ILD.

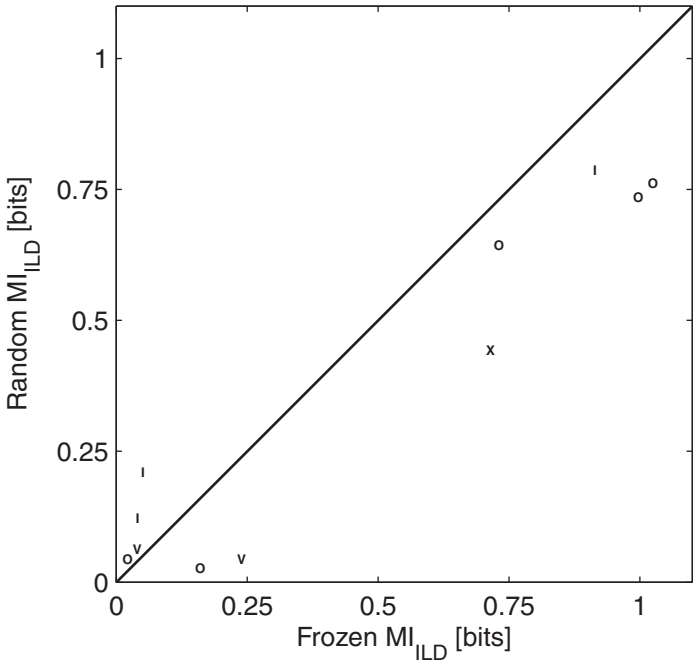


Figure 6.6: MI carried about ILD cues in the random noise stimulus set plotted as a function of the MI carried about ILD cues for frozen noise stimuli.

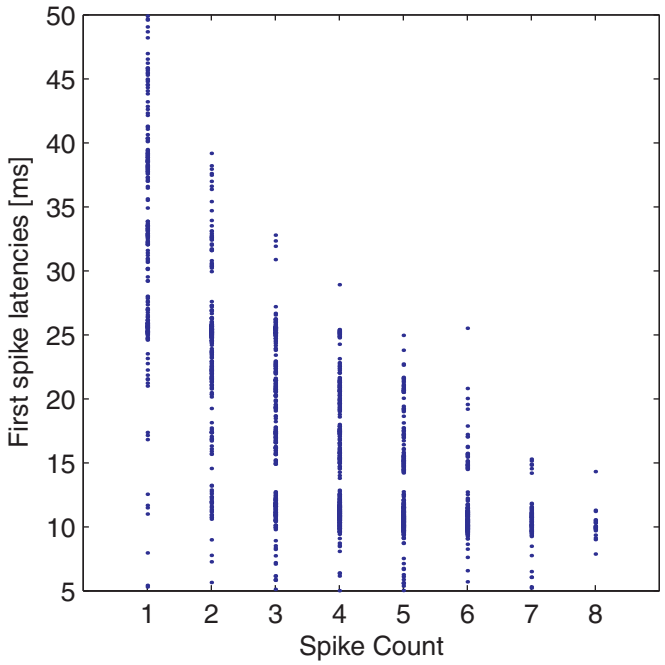


Figure 6.7: Correlation between first spike latency and spike count. Here, the first spike latencies are plotted as a function of the number of spikes within the analysis window (5 to 50 ms), for one example neuron in response to the ABI/ILD stimulus set.

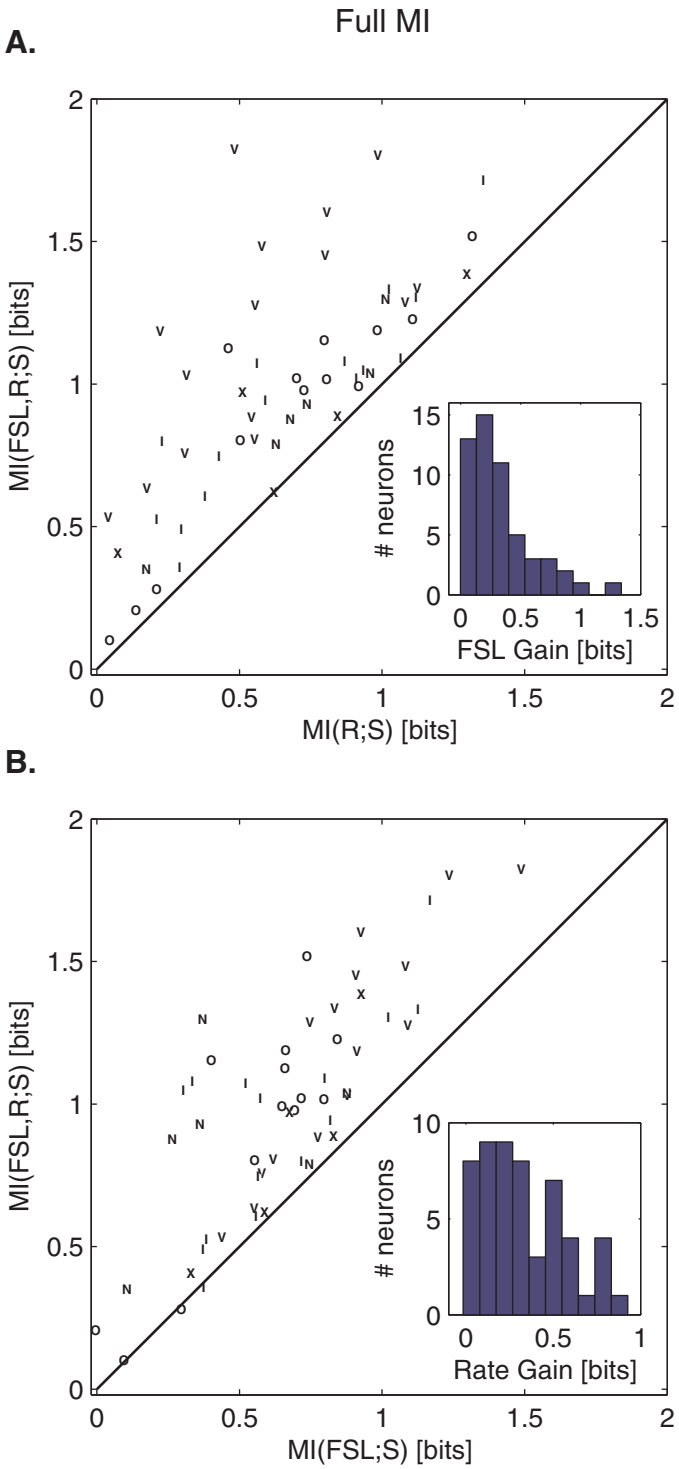


Figure 6.8: Joint rate/FSL information. **A.** The MI in a joint rate/FSL code is plotted as a function of the information conveyed by a rate code alone. *Inset:* histogram of each neuron's extra information in FSL not conveyed by rate. **B.** Joint rate/FSL MI plotted as a function of the information conveyed by a first spike latency code alone. *Inset:* extra information in rate not conveyed by first spike latency.

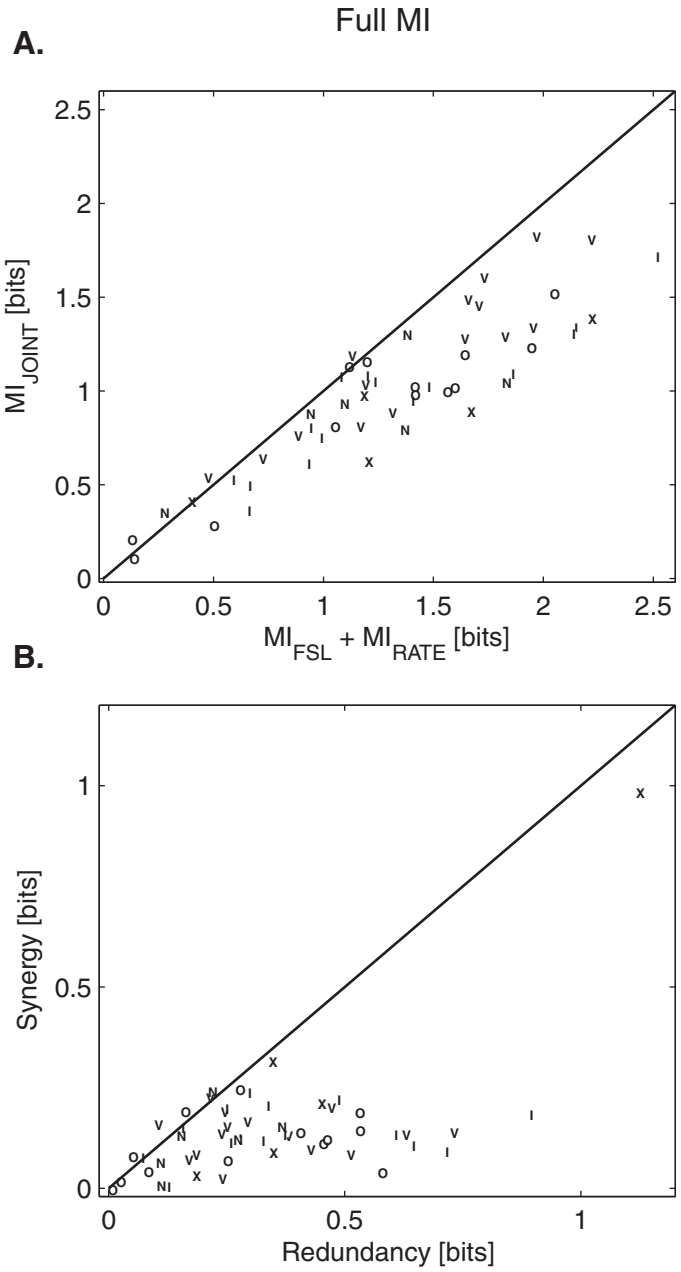


Figure 6.9: Synergy and redundancy in the joint rate/first spike latency code. **A.** MI_{JOINT} is plotted as a function of the sum of the MIs in the individual codes. **B.** The synergy, MI(R;FSL|S), is plotted as a function of the redundancy, MI(R;FSL). (The outlier is a peculiar neuron that, at the start of recording had a clear type-I response map, but adapted gradually to an onset response. It is the only such unit in the population.)

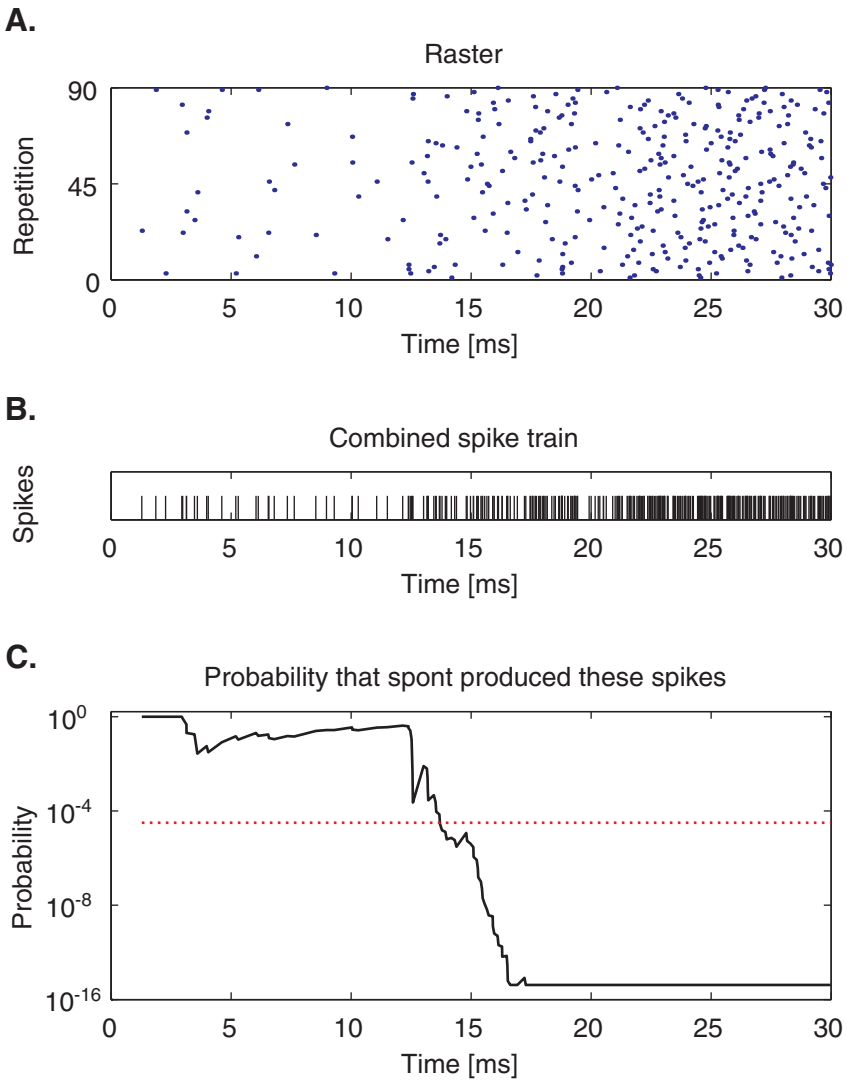


Figure 6.10: Detecting first spike latency in the presence of spontaneous discharges. **A.** Spike raster plot of the first 30 ms of one neuron's response to one stimulus. **B.** The spikes from A are combined into a single spike train. **C.** For each spike in train B, the probability that the spontaneous rate would produce a response at least as large as the last n spikes in the observed time is computed for n ranging from 5 to the number of spikes observed so far. The minimum of these probabilities is plotted as a function of spike time. The latency is the first time at which this probability exceeds threshold (dotted red line).

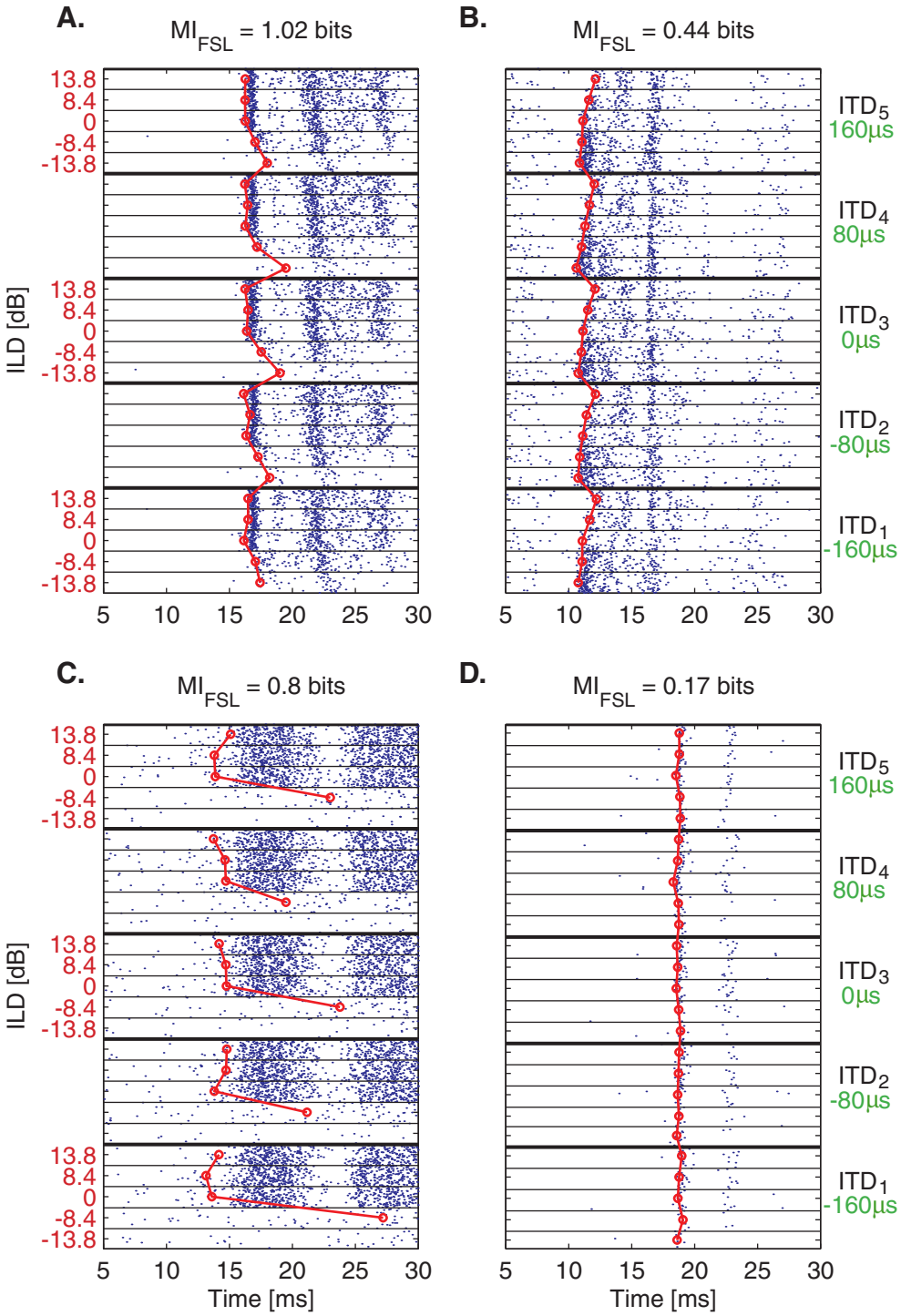


Figure 6.11: First spike latency functions from four example neurons responding to the ITD/ILD stimulus set. Red lines give the output of the first spike latency dectection algorithm.

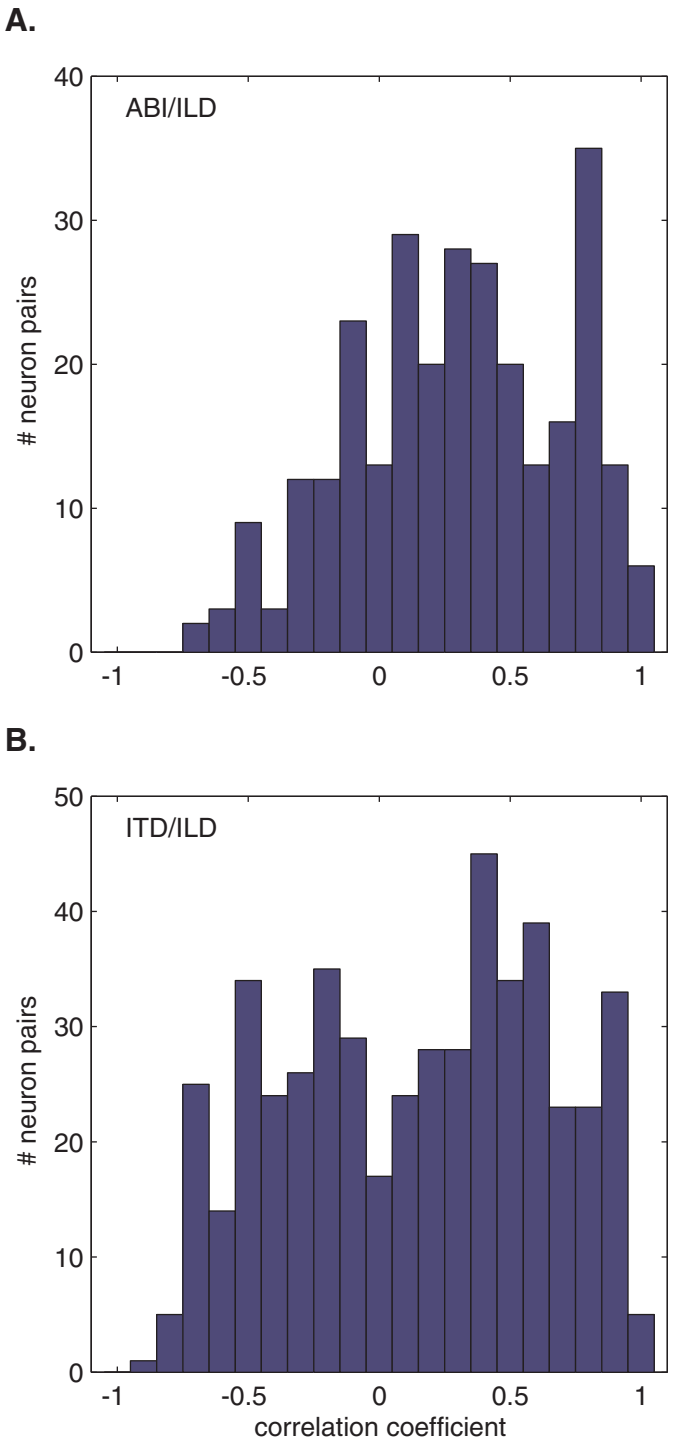


Figure 6.12: Histograms of pairwise correlation coefficients between first spike latency functions. **A.** ABI/ILD stimulus set. **B.** ITD/ILD stimulus set.

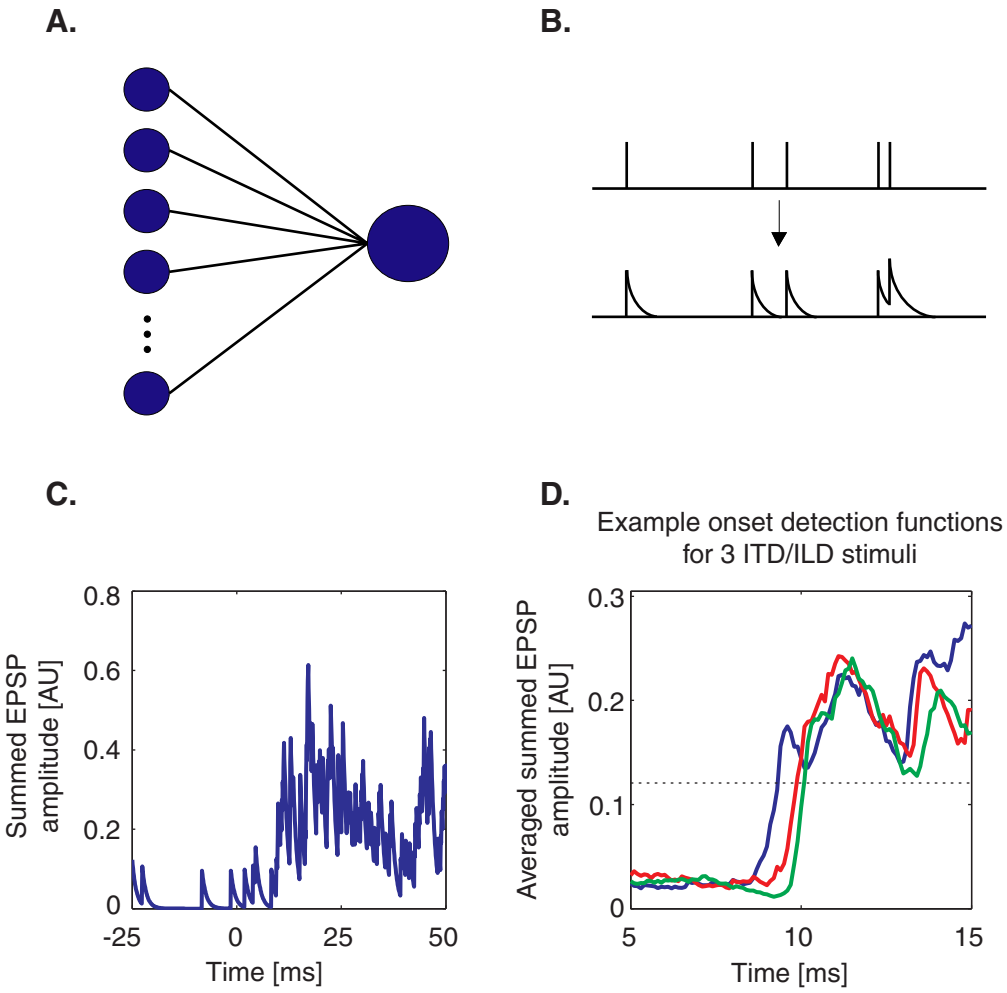


Figure 6.13: The coincidence detector model. **A.** Schematic of the model, where one coincidence detector neuron receives input from every cell in the population. **B.** Every spike from the pre-synaptic neurons is convolved with an exponential tail (with 1 ms decay time) to mimic the EPSP it would induce on the coincidence detector. **C.** The summed EPSP amplitude from one run randomly chosen set of spike trains from one stimulus in the ITD/ILD stimulus set. **D.** The average summed EPSP amplitude from 500 randomly chosen sets of spike train inputs to the coincidence detector. Traces from three different ITD/ILD stimuli are shown. The dotted line represents the threshold for stimulus onset detection.

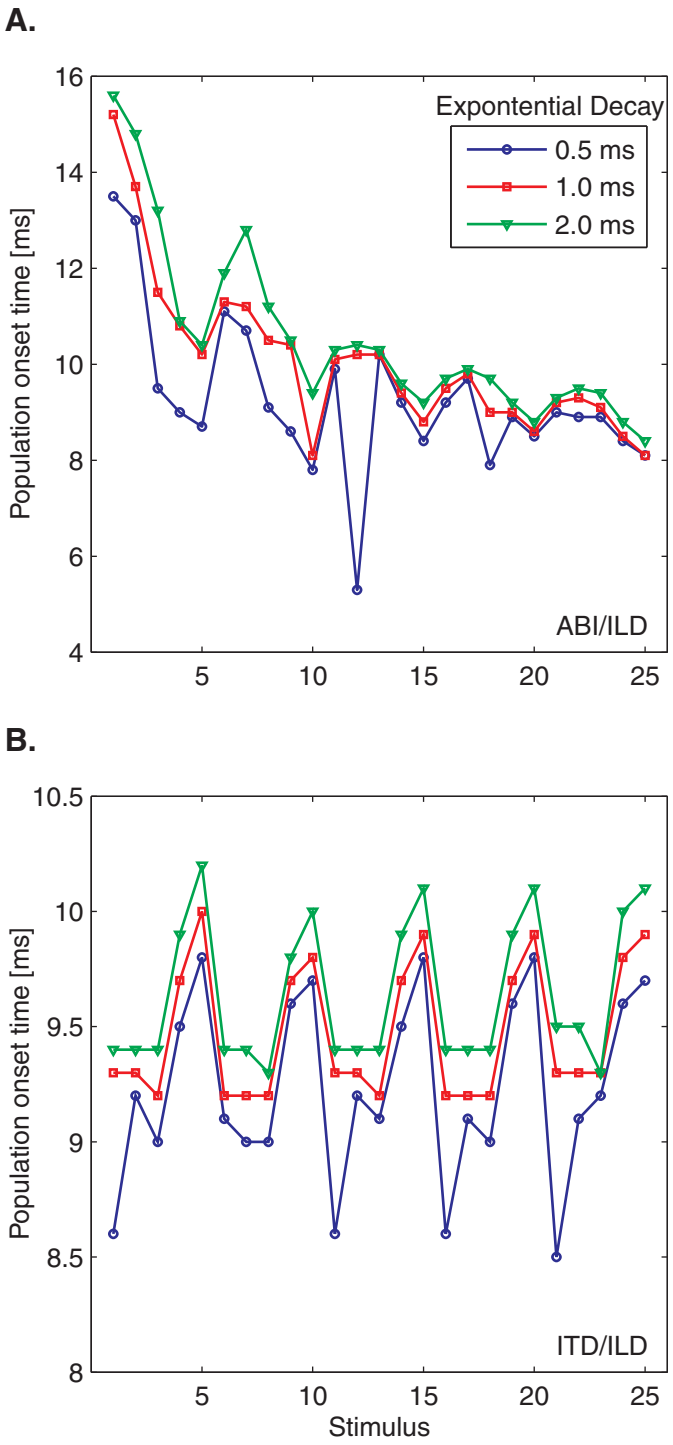


Figure 6.14: Population-derived stimulus onset times from the coincidence detector model, using 3 differrent values of the exponential decay constant. **A.** Onset times for the ABI/ILD stimulus set. **B.** Same, for the ITD/ILD stimulus set.

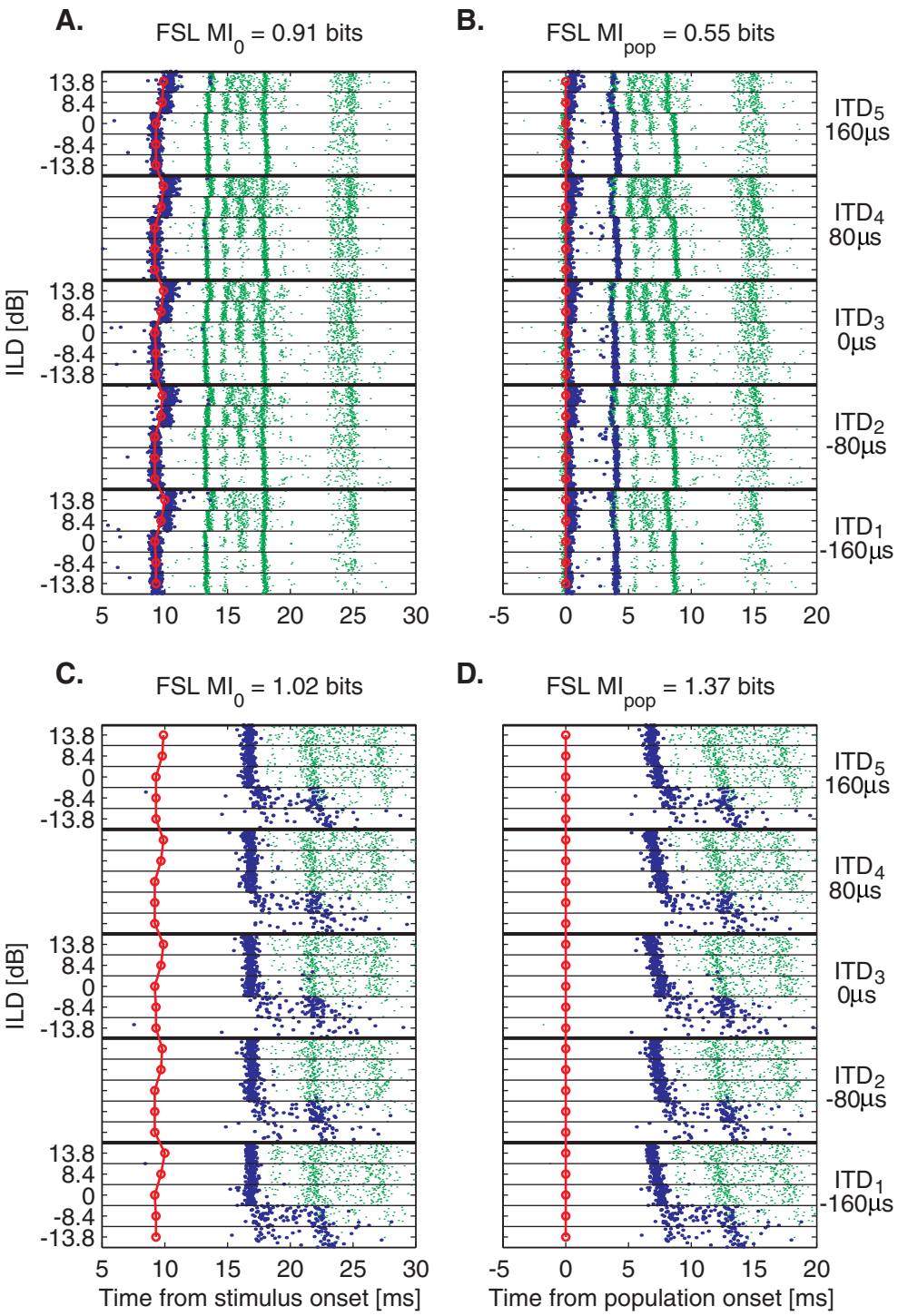


Figure 6.15: Measuring first spike latency relative to the population onset. **A.** Spike raster plot of the first 30 ms of one neuron's response to the ITD/ILD stimulus set. All spikes are shown as small green dots, the first spikes within the [5,50] ms window are overlaid with a larger blue dot. The population-derived onset time is shown in red. **B.** Same rasters as in A, where the spike times are given relative to the population-derived stimulus onset. The blue dots represent the first spikes in a 45 ms window starting at the population onset. The full MIs for each condition are shown above the plots. **C, D.** Same as A and B, for a different example neuron.

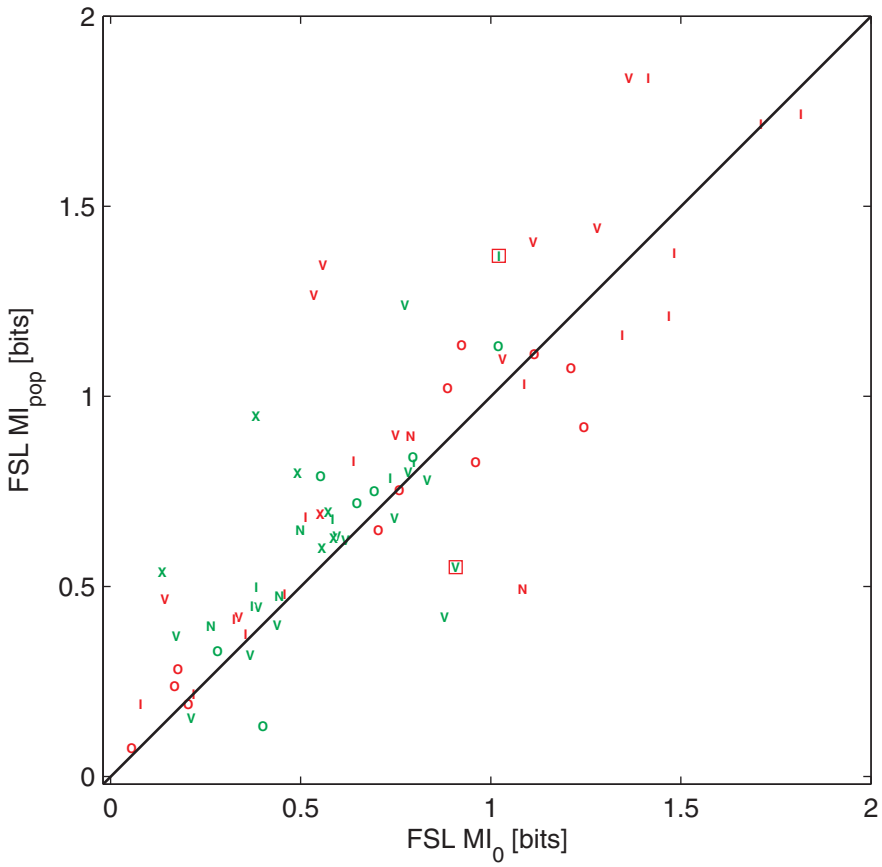


Figure 6.16: A comparison of the FSL MIs measured relative to an assumed zero time reference (MI_0) and relative to the population-derived onset time (MI_{pop}). The example neurons from Fig. 6.14 are shown surrounded by red squares. Color denotes the stimulus set, with green for ITD/ILD and red for ABI/ILD.

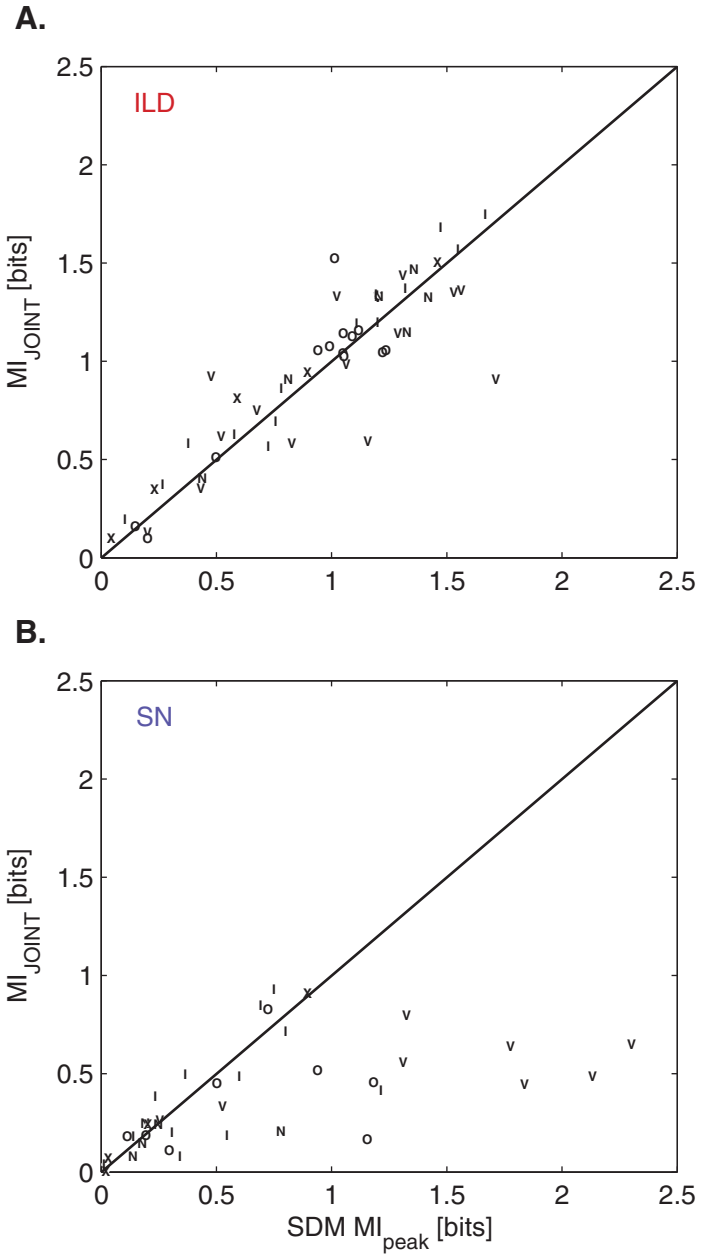


Figure 6.17: Comparison between the information in a joint rate/first spike latency code and the peak informatoin recovered by an SDM decoder, for an analysis window extending from stimulus onset to 20 ms past stimulus offset. **A.** ILD information. The joint code captures as much as the SDM decoder. **B.** SN information. The joint code consistently underestimates the information.

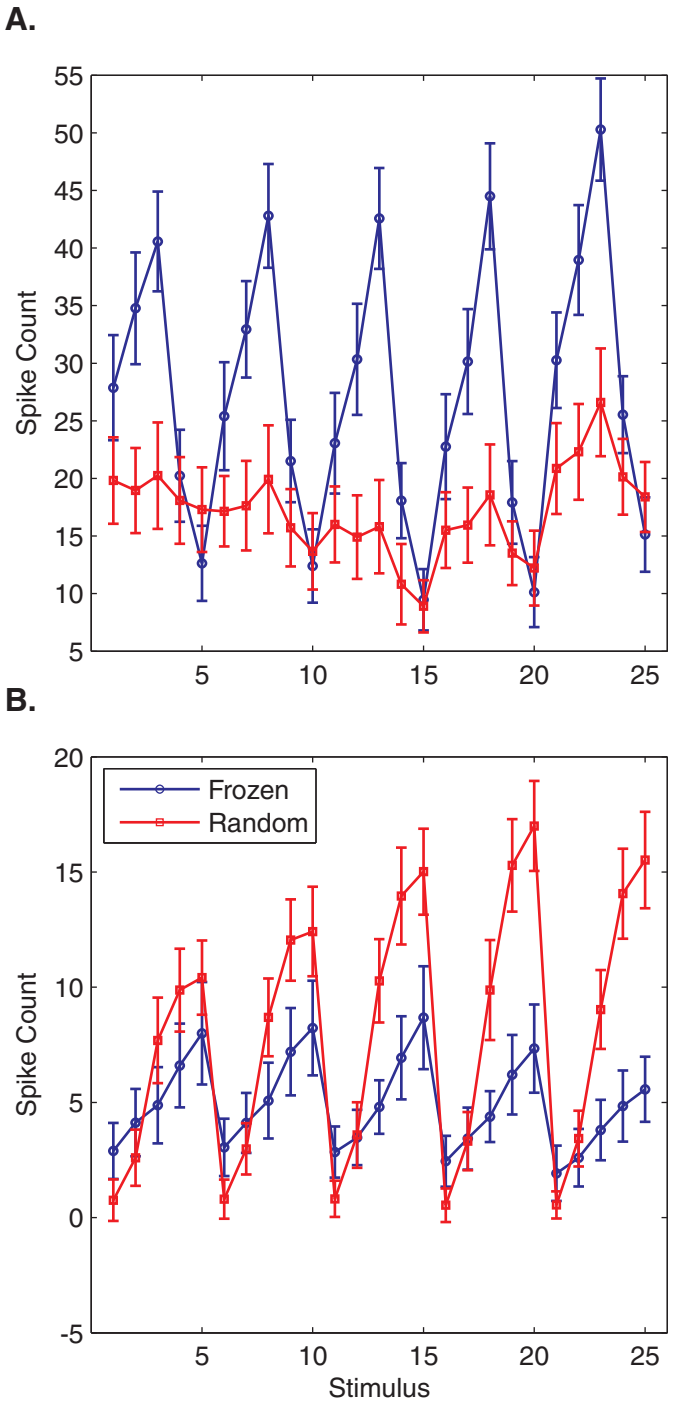


Figure 7.1: Comparison between correlated frozen and uncorrelated random noise. Each plot shows the mean and standard deviation of the spike counts in response to either the frozen or the random noise. ILD and SN cues are identical between the two stimulus sets. The numbering is as follows: in 1-5, the ILD increases from ipsilateral to contralateral while SN is held constant, numbers 6-10 have the same range of ILDs for another SN, etc. **A.** 5.5 kHz neuron, though with a secondary excitatory region at 19kHz. Its response is more to ABI than ILD, as it peaks around 0 azimuth, where the level of these stimuli is highest. **B.** 3 kHz type-I neuron.

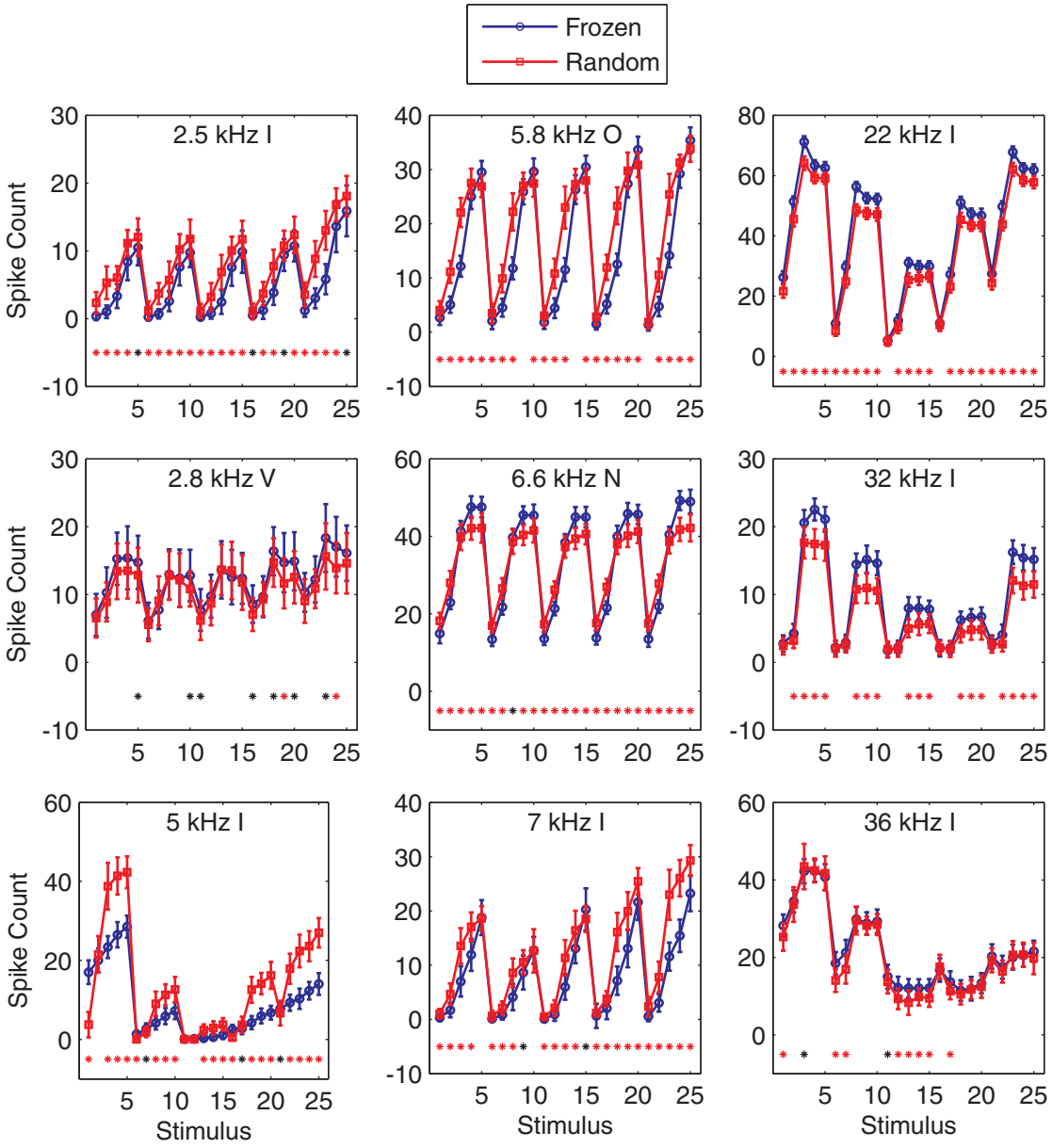


Figure 7.2: Comparison of correlated frozen noise and uncorrelated random noise for the nine other neurons recored with these stimuli. Black asterisks on the bottom responses that are significantly different at the p<.05 level, red asterisks indicate a p<.01 significance. BF and cell type are printed in the center of each plot, BFs are organized by column into roughly similar frequency ranges.

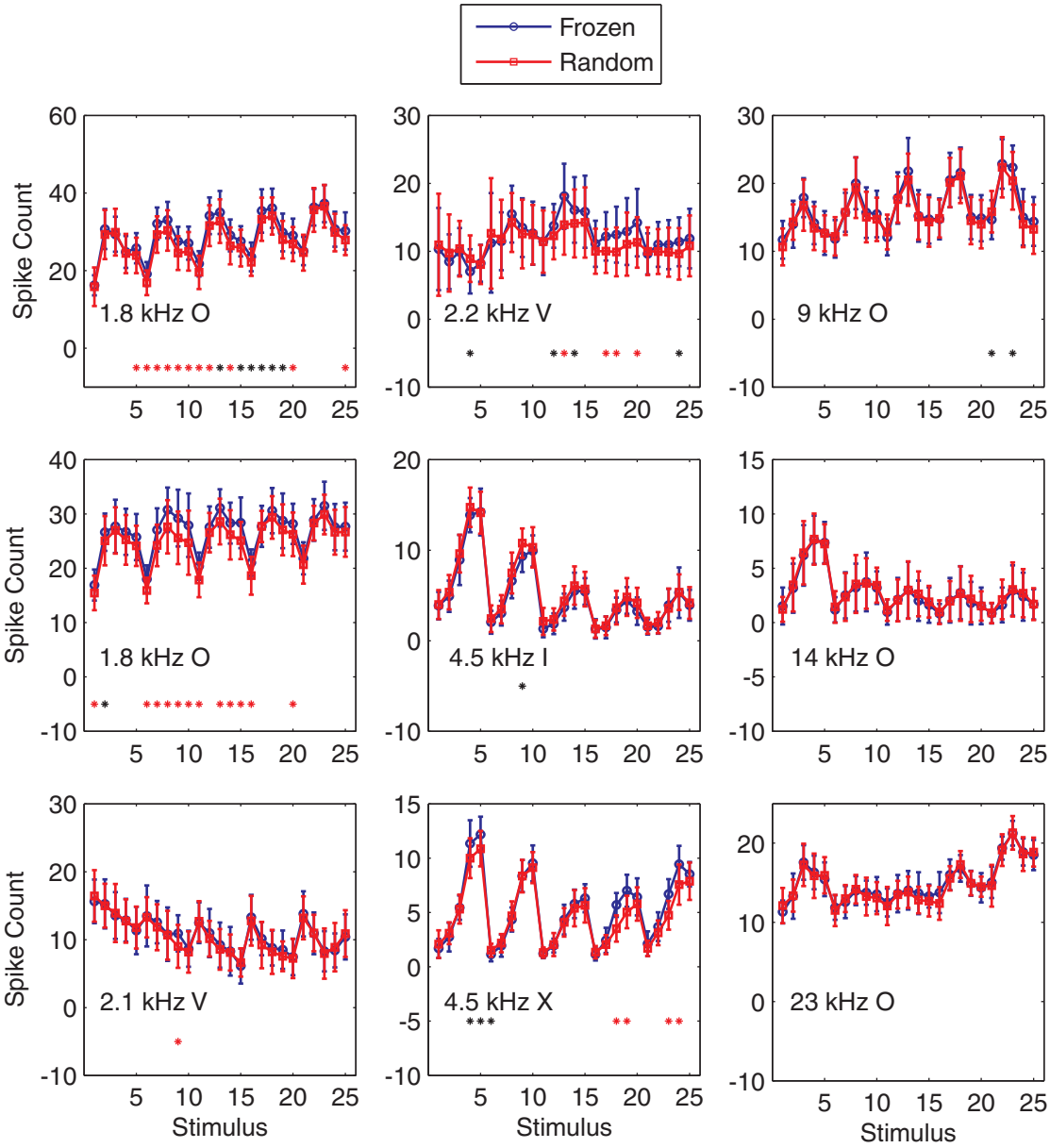


Figure 7.3: Comparison of correlated frozen noise and correlated random noise. Format is the same as the last figure.

Biography

Steven Chase was born to parents Kathleen and Michael, on October 21, 1974. His birth brought the population of his hometown, Mason, NH, up to 835.

His decision to go to college was made after a particularly difficult night working as a night-shift janitor at a local supermarket. He attended Caltech from 1993 to 1997, earning his B.S. in Applied Physics while doing research on laser-induced fluorescence. Much to the disgust of his advisor, who wanted him to be a physicist, he joined the Ph.D. program in Electrical Engineering at UC Berkeley in the fall of 1997 to design fiber optic lasers.

While preparing for a journal club one day (researching how the retina compared to silicon photodetector technology), he began wondering why one could not build an artificial eye. Quickly realizing that it was because you couldn't plug it in, he started looking into nerve-electrical interfaces, neural prosthetics, and sensory coding.

His new interest wouldn't wait. He left UC Berkeley early, with an M.S. degree, and proceeded to the Johns Hopkins University in the fall of 1999 to study Biomedical Engineering. He has never regretted the decision.

He joined the lab of Eric Young in the spring of 2000 to work on the application of information theory to auditory sensory processing. His post-doctoral work will continue the study of how single neurons represent the world, by studying how neurons in the motor cortex adapt and learn to use neural prosthetic devices.